UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Essays on Mutual Fund Strategies and Investor Characteristics

Autor:

Hang Dong

Directores:

Javier Gil-Bazo

Pablo Ruiz-Verdú

Doctorado en Economía de la Empresa y Métodos Cuantitativos

Departamento de Economía de la Empresa

Mayo de 2014

TESIS DOCTORAL

ESSAYS ON MUTUAL FUND STRATEGIES AND INVESTOR CHARACTERISTICS

Autor: **Hang Dong**

Directores: Javier Gil-Bazo
Pablo Ruiz-Verdú

Firma del Tribunal Calificador:

	Nombre y Apellidos	Firma
Presidente		
Vocal		
Secretario		

Calificación:

Getafe, de de 2014

Acknowledgements

It is a great pleasure for me to express my gratitude to the people who have been an important part of my life during the past few years.

First of all, I am indebted to my supervisors, Javier Gil-Bazo and Pablo Ruiz-Verdú, for all of the supervision, their selfless time, their insightful and useful comments and supporting me all the way from zero to the PhD degree. Without them, this thesis would have been impossible.

I am also thankful to the Business Department at the University Carlos III of Madrid, from which I was funded. I have spent a great time here.

A warm thanks goes to all of the nice people I have met at the University Carlos III of Madrid. I have experienced a colorful life in Madrid. Without them, it would would be unbearably tedious.

Finally, I am forever thankful to my parents for all of their love and support. 我爱你们，爸爸妈妈！As the words to my future partner and children: You could have appeared in my life at least 3 years earlier without this thesis.

Contents

Resumen:

Esta tesis estudia las estrategias usadas por los fondos de inversión para atraer inversores, así como las características de éstos asociadas a cada estrategia. Específicamente, examinaré la política de dividendos y la heterogeneidad en en características de los inversores a través de los rendimientos esperados y la política de dividendos. El primer capítulo se centra en cómo los fondos de inversión determinan su política de dividendos, i.e. frecuencia con la que éstos se pagan. Considero que, para determinar su política de dividendos, los fondos necesitan solucionar dos conflictos de intereses: entre nuevos y antiguos inversores, y entre los inversores existentes. La evidencia empírica que presento apoya mi argumentación. El segundo capítulo investiga si algunos grupos de inversores son más propensos a cometer errores (e.g. escogiendo fondos con comisiones más altas) que otros. Proponemos la utilización de nuevos datos como las características socio-demográficas de los visitantes a las páginas web de los fondos , como una aproximación a las características de los inversores en esos fondos. Los resultados son consistentes con la literatura previa en cuanto a que inversores de mayor edad y con bajos ingresos son más propensos a comprar fondos con unas características concretas de las cuales se espera que tendrá un rendimiento inferior en el futuro. El tercer capítulo intenta aportar nueva evidencia sobre la existencia de clientelas de dividendos en el mercado de fondos de inversión. Demuestro que inversores de mayor edad y menores ingresos tienden a comprar fondos que reparten dividendos con mayor frecuencia con más probabilidad.

Abstract:

This thesis studies the strategies that mutual funds use to attract individual investors and the investor characteristics that are associated with the strategies. More specifically, I examine the influence of fund strategies, particularly dividend policy, on investors and how investors react differently to fund strategies and performance depending on their socio-demographic characteristics. The first chapter addresses the question as to how a mutual fund determines its dividend policy, i.e., the frequency of dividend distributions. I argue that in determining their dividend policy, mutual funds need to solve two conflicts of interest: between new and existing investors and within existing investors. I present empirical evidence that supports the theoretical arguments. The second paper investigates the question of whether some groups of investors are more prone to making mistakes (e.g., choosing high-fee funds) than others. I propose a new data set on the socio-demographic characteristics of visitors to mutual fund websites as a proxy for the characteristics of people investing in those mutual funds. I find results consistent with previous literature in that old and low-income investors are more likely to buy mutual funds that are predicted to underperform based on their characteristics. The third chapter provides direct evidence on the existence of dividend clienteles in the mutual fund market. I document that old and low income investors are more likely to purchase funds with high dividend frequency.

1

Introduction

Over the past decade, the assets managed by mutual funds have increased dramatically in the US. The assets under management in the US mutual fund industry grew from $3,526 billion at the beginning of 1998 to $13,045 billion at the end of 2012.[1] The nearly fourfold increase stems from 1) the increase of equity market; and more interestingly, 2) the net new cash flow, i.e., the dollar value of new fund sales less redemptions. The Investment Company Institute (ICI) reports that from 2002 to 2012, individual investors (households) invested an average of $349 billion each year, on net, in long-term registered investment companies, mainly mutual funds. And 44.4 percent of households invest in mutual funds. Thus, ICI concludes that millions of individual investors have increasingly turned to mutual funds to achieve their financial goals. The purpose of this thesis is to investigate how mutual fund strategies affect investors and how investors react to those strategies, as well as the effect of investor characteristics on the purchase decisions. More specifically, I examine the influence of fund strategies, particularly dividend policy, on investors and how investors react differently to fund strategies and performance depending on their socio-demographic characteristics.

The compensation for mutual fund management companies is mostly derived from management fees, charged in the form of a percentage of the assets under management. To maximize their revenue, mutual funds compete with others to attract investors. The most important factor in this competition is performance. Carhart (1997) and Sirri and Tufano (1998) are among the first to study the relationship between past performance and net fund flows. They find that funds with superior past performance are associated with disproportionately large new money inflows, while poor-performing funds do not suffer large outflows. The convexity of the relationship between performance and new cash flow may create the incentives for mutual funds to alter the riskiness of the portfolios in order to maximize the inflows (Brown et al., 1996; Chevalier and Ellison, 1997).

Funds also use other strategies to attract new flows. For example, Jain and Wu (2000) and Gallaher et al. (2006) show that advertising has significantly positive effects on mutual fund inflows, which suggests that individual investors follow the advertising to buy mutual funds. Mutual funds also set fees and distribution channels in order to target different types of investors (Bergstresser et al., 2009; Christoffersen and Musto, 2002; Gil-Bazo and Ruiz-Verdú, 2009). Barclay et al. (1998) show that mutual funds trade off the welfare of their existing and new shareholders in choosing the realization policy of unrealized capital gains. Christoffersen et al. (2005) find that mutual funds make trade offs within retirement and nonretirment accounts to cater to their shareholders. More recently, Harris et al. (2012) provide evidence on that some mutual

[1] All statistics in Introduction are from 2013 Investment Company Fact Book, ICI

funds purchase high dividend yield stocks before dividend payment dates for the purpose of increasing dividend incomes, and therefore, receiving larger inflows.

Despite the positive association between flows and fund strategies, any given strategy is not likely to be attractive to all investors. For example, though funds could maximize the expected flows by changing the riskiness of the portfolios, it may drive away some investors who are sensitive to the risk. Barclay et al. (1998) and Christoffersen et al. (2005) provide clear evidence that funds' strategies sometimes hurts some investors.

It is not always easy to determine who are the investors benefited or hurt by a certain fund strategy. A clear difference can be made between new and existing investors, as in Barclay et al. (1998). Mutual funds are supposed to act in behalf of the existing investors. Yet, some funds might favor new investors in order to get more inflows. Investors also differ in their socio-demographic characteristics. Previous studies show that investor preferences, the degree of sophistication, and thus, behaviors vary across investor characteristics. For instance, Miller and Modigliani (1961) conjecture that investors with different characteristics do not have the same attitude toward dividends and, therefore, behave differently. Graham and Kumar (2006) and Becker et al. (2011) find strong empirical evidence supporting this conjecture. Previous literature also shows that the fees and performance of mutual funds are different across investor socio-demographic characteristics.[2]

The primary objective of this thesis is to study the interrelation between fund strategies and investor characteristics. The thesis is structured in three chapters. Chapter 2 of this thesis examines mutual funds' choice of dividend payment frequency, a fund strategy that has been neglected by the literature. Mutual funds' dividend payment is not as flexible as firms', because regulation essentially requires mutual funds to pay out nearly all dividends each year. However, mutual funds can still decide the frequency of dividend distributions. This chapter argues that multi-dividend policy hurts mutual fund shareholders and investigates the reasons why some mutual funds distribute dividends more frequently than mandated. I propose that funds deliberately set dividend policy to increase assets under management and, thus, fee revenues. In determining their dividend policy, mutual funds need to solve two conflicts of interest: (1) between new and existing investors, and (2) within existing investors with different dividend preferences. The empirical results show that the probability of a fund choosing a multi-dividend policy is associated with fund characteristics that affect both conflicts of interest. We also

[2]See, for example, Bailey et al. (2011); Engström (2007); Grinblatt et al. (2011); Niebling et al. (2009); Tang et al. (2010).

find evidence that multi-dividend policy alters the sensitivity of mutual fund net flows to the fund characteristics that affect both conflicts of interest.

Chapter 3 investigates an important unresolved question: Who buys those funds that are expected to underperform? The question is raised by the abundant empirical evidence on the persistence of mutual fund underperformance. However, attempts to answer this question have been hampered by lack of comprehensive data on individual investor decisions in large markets. To overcome this data limitation, we use the sociodemographic characteristics of US visitors to mutual fund websites to proxy for investor characteristics. We find that differences in sociodemographic characteristics are systematically associated with predicted fund performance: Funds with a higher fraction of female, older or low-income investors are associated with worse predicted performance. We also find that differences in the optimality of investor choices across investor sociodemographic groups can be explained by differences in sensitivity to past performance and fund fees. Finally, there is limited evidence that fund marketing can explain why some groups of investors buy the underperforming funds.

Chapter 4 addresses another interesting question: Who is the dividend clientele in mutual funds? Previous studies show that there is a dividend clientele in the stock market by identifying its characteristics (Becker et al., 2011; Graham and Kumar, 2006). The evidence is consistent with the conjecture that older and low income investors have a preference for the dividends. However, there is no such study for the mutual fund market. This study is important for us to better understand the mutual fund market and protect invulnerable investors. Using the data employed in Chapter 3, I find limited evidence supporting that older and low income investors buy the funds with more frequent dividend distribution. The study sheds light on the existence of dividend clienteles and helps explain the heterogeneity of mutual fund dividend policies.

The thesis contributes to existing research in three ways. First, I document a new fund strategy, i.e., dividend policy, that has been ignored by prior literature, but is important to investors. Second, I explore a new Internet database, which can be used as a proxy for investor characteristics of all US investors. Such a large scale database with the characteristics of US investors was not previously available. Third, I provide evidence of dividend clienteles in the mutual fund market.

2

Mutual Fund Dividend Policy

Contents

2.1 Introduction and Literature Review

In the US, a mutual fund[1] is defined as a regulated investment company in the US.[2] As such, a mutual fund pays dividends to its shareholders, like other types of corporates.[3] However, mutual fund dividend policy is different from the dividend policy of corporates in many aspects, such as its sources, distributions, and regulations. Therefore, it is surprising that mutual fund dividend policy receives no attention in contrast to the large body of literature on corporate dividend policy. This chapter investigates how a mutual fund determines its dividend policy and its impact on investors.

One possible reason for the lack of academic attention to mutual fund dividend policy is that a mutual fund has a less flexible dividend policy compared with a corporate. A corporate determines its dividend policy in terms of "how much, when and how" during the life of the corporate (DeAngelo et al., 2008). However, a mutual fund, as described below, is highly regulated in all aspects and has to pay out all dividends in the form of cash each year. Yet, the rule does not restrict the frequency of dividend payout. A mutual fund could pay dividends, if there is any, once or more times per year to mutual fund investors. Therefore, a mutual fund can still have a dividend policy, namely, the number of times it pays dividends during the year.

Dividend policy varies across mutual funds. In my sample, 30% of funds choose a multi-dividend policy, i.e., pay dividends more than once during a year. Intuitively, a multi-dividend policy is costly for both investors and mutual funds. For example, there are transaction costs for mutual funds to pay dividends. Some of the costs (e.g., transference fees) might be fixed for each distribution. Therefore, a multi-dividend fund may incur higher transaction costs, which are ultimately paid by investors. Multi-dividend funds also lose fee revenue on dividends that not automatically reinvested before the end of the year. These costs raise the question: Why do many funds choose to pay dividends more often than mandated?

One possible answer is that mutual funds could attract inflows to the funds and, therefore, increase fee revenue by deliberately setting dividend policy. This chapter proposes that dividend policy results in two conflicts of interest: between existing and new investors, and within existing investors with different dividend preferences. As a consequence, it is reasonable to suspect that mutual

[1] Throughout the chapter, we use "mutual funds" or "funds" to refer to retail equity mutual funds unless otherwise specified.

[2] See http://www.sec.gov/answers/mfinvco.htm and http://www.sec.gov/answers/mutfund.htm.

[3] We use corporate to refer to the company other than regulated investment company.

funds trade off the interests of different investors. This chapter asks three questions: Does multi-dividend policy hurt existing investors? Do mutual funds choose their dividend policy to trade off interests of different investors? Do investors respond to dividend policy?

Retail investors have to pay income taxes on dividends received in the current year. New investors would prefer to purchase funds with less undistributed dividends and, therefore, pay less taxes. To attract new investors, some mutual funds might pay dividends more frequently than it is legally required to lower the tax burden for new investors. However, this policy might hurt existing investors who, on the contrary, prefer single-dividend policy. Therefore, mutual funds must trade off the interests of existing and new investors.

The trade-off aforementioned extends the previous literature on mutual funds' agency problems. The chapter shows evidence that a mutual fund might choose the optimal dividend policy to maximize its flow and revenue at the cost of investors. A mutual fund management company's compensation is mostly derived from management fees, in the form of a percentage of assets under management. Mutual funds use two ways to maximize their interest: attracting more flows and/or increasing the fees. For example, Chevalier and Ellison (1997) and Busse (2001) show that fund managers alter the riskiness of portfolios depending on the fund's past performance. This operation might increase the expected inflow because of the convex shape of the relationship between performance and flows. Barclay et al. (1998) show that mutual funds trade off the welfare of their existing and new shareholders in choosing the realization policy of unrealized capital gains. These authors provide evidence of an agency problem resulting from the demand for new investment by mutual fund managers. To attract new fund inflows, fund managers may realize and pass through unrealized capital gains and reduce the tax overhang caused by unrealized gains at the cost of existing investors. More recently, Harris et al. (2012) show that some mutual funds purchase high dividend yield stocks before dividend payment dates for the purpose of increasing dividend incomes, and therefore, receiving larger inflows. Mutual funds also trade off the interest of different types of investors by setting fees. Christoffersen and Musto (2002) argue that the pricing of mutual funds depends on their demand curves. As such, mutual funds might charge high fees to certain types of investors. Gil-Bazo and Ruiz-Verdú (2008) theoretically show that some funds might charge higher fees to unsophisticated investors in equilibrium. Consistently with their prediction, Gil-Bazo and Ruiz-Verdú (2009) find a negative relation between fees and before-fee performance in the cross section of US equity funds. Moreover, Bergstresser et al. (2009) document little or no benefit from brokers when investors are charged large distribution fees.

This chapter also extends the literature that relates the conflicting preferences of existing investors and mutual fund strategy. More specifically, I argue that, when choosing their dividend policy, mutual funds need to resolve a conflict between existing investors with different dividend preferences. Existing investors have different interests. Johnson (2004) documents that the transaction costs of investors depend on their investment horizons. Short-term investors transfer transaction costs to long-term investors. Christoffersen et al. (2005) show that retirement and nonretirment accounts have different tax preferences. Consequently, mutual funds make trade offs between the two types of accounts. More related to our topic, differences in dividend preferences across investors are well documented in the literature (see, for example, Becker et al., 2011; Graham and Kumar, 2006; Scholz, 1992; Thaler and Shefrin, 1981). One explanation of the existence of dividend clientele is that some investors may use regular stock dividend income streams to finance consumption (Becker et al., 2011; Graham and Kumar, 2006). If this argument is true, these investors prefer not only high dividend assets but also more frequent distributions.

The empirical analysis in this chapter proceeds in two parts. In the main test, I investigate how dividend policy is related to fund characteristics that affect the different conflicts of interest. In the second test, I examine how investors respond to multi-dividend policy. Specifically, I regress net relative flows on fund characteristics and add interaction terms of fund characteristics and dummy for multi-dividend policy.

The main results are summarized as follows. First, I find evidence that mutual funds deliberately choose dividend policy. The dividend policy is persistent over time, and, to large extent, independent from market characteristics, in particular of the market average stock dividend ratio. Second, I find that multi-dividend policy potentially hurts investors. Multi-dividend funds on average have more stable dividend payout ratios over time and obtain worse raw returns than single-dividend funds. The results suggest that multi-dividend funds chase dividends and hold liquid assets. Third, I document that the frequency of dividend payment is positively associated with the dividend ratio, a variable that affects the conflict of interest between new and existing investors. I also find that participation costs, proxied by marketing expenses, are positively related with dividend frequency. Funds are more likely to pay dividends more frequently in the years when the risk-free interesting rates decrease. Finally, I show evidence that funds with higher dividend ratios and participation costs attract more net flows if they use multi-dividend policy, other things equal. The investors in multi-dividend funds are less sensitive to high performance than the ones in single-dividend funds. The results are consistent with the existence of a dividend clientele, who uses dividends to finance their consumption.

The rest of the article is organized as follows. Section 2.2 covers the relevant background. Section 2.3 provides the hypotheses. Section 3.2 describes the data. Section 2.5 shows the empirical results. Finally, section 3.4 concludes.

2.2 Background

Mutual funds are categorized as regulated investment companies in the US. Under the tax rules in subchapter M of the Internal Revenue Code (IRC), mutual funds do not pay taxes on their income, i.e., dividend or interest received from asset under management, or capital gains as long as they fulfill certain requirements. One requirement on distribution is that a mutual fund should "pass through" at least 90% of incomes to their shareholders in the US each taxable year. The mutual fund may retain up to 10% of its incomes and all capital gains, which are taxed at regulated corporate tax rate. The IRC also imposes an excise tax, at 4% rate, on mutual funds unless a mutual fund distributes 98% of its ordinary income during the calendar year before December 31 and 98% of capital gains earned in the 12-month period ending on October 31.[4] Therefore, mutual funds usually distribute nearly all dividends and realized capital gains each year to avoid unnecessary taxes.[5]

It is worthy to mention that dividend received by mutual funds and dividend received by fund shareholders are different in terms of quantity. Although mutual funds have to "pass through" dividends to their shareholders, it does not necessarily follow that shareholders receive all dividends that a mutual fund receives from its assets. A fund uses dividends received from assets under management to offset expenses. As a result, the dividend received by shareholders is the dividend received by the fund from its assets under management net of the expense. Thus, in rare cases, realized dividend frequency is not equal to scheduled frequency if the expense a fund charges is higher than the dividend it is supposed to pass through.

In this chapter, we only focus on the dividend and its policy. Mutual funds typically generate three types of current of potential cash flow for investors who do not sell their shares. These are the following: (1) incomes, i.e., dividends payments, including income in the form of dividends and interest on the securities in its portfolio (minus disclosed expenses)[6]; (2) realized capital gains, i.e., appreciation of securities in value, which are already sold minus any capital losses (it includes short-term realized capital gains and long-term realized

[4]See Section 852 in Subchapter M of the IRC for more details.

[5]See the chapter "Tax Features of Mutual Funds", Investment Company Fact Book, ICI, 2012, for more details.

[6]www.sec.gov/investor/pubs/sec-guide-to-mutual-funds.pdf.

capital gains); (3) unrealized capital gains, i.e., appreciation of a fund's assets in value, which are not sold yet by fund managers. The virtue of addressing mutual fund dividends rather than capital gains (realized or unrealized) is obvious. First, a fund's dividend yields are highly representative of the fund's ex ante dividend policy because dividends are predicable. Meanwhile, a fund's capital gains and their realization are highly dependent not only on ex ante policy but also on other market factors, such as stock prices and fund flows (Christoffersen et al., 2005). Second, the dividend distribution is observable for investors from the prospectus and past fund distributions, but unrealized capital gain is not. Third, dividends have more direct and large impact on investors than capital gains (for more details, see Barclay et al. (1998) and section 2.3.2). Fourth, although dividends and realized capital gains have the same framework and impact on investors, dividend policy is more heterogeneous than capital gain policy across funds because mutual funds typically save realized capital gains until the end of the year because it could be used to offset capital loss, but not for dividends.

We define dividend policy as the frequency of dividend distributions that a mutual fund intends to pay out during one year. This definition is similar to, if any, dividend distribution schedules in their prospectuses. For example, the Fidelity Equity-Income Fund claims that its dividend distributions are in April, July, October, and December.[7] The scheduled frequency of the Fidelity Equity-Income Fund is four times per year. It could be seen as the promise to investors. I verify that a mutual fund should strictly follow the schedule in its prospectus, if any, to pay dividends. As the consumer service of Fidelity Investments replies to my request, "within a mutual fund's composition, cash is set aside to pay dividends to shareholders of the security. If the fund is scheduled to pay a distribution on a quarterly basis, it will pay one as long as enough cash within the fund is available." There is no reason to suspect that a fund documents a multi-dividend policy in prospectus and it actually intends to apply single-dividend policy. It is costly for the fund if such behavior is found, and investors who purchase this fund because of dividend policy might flee from the fund when they find the realized dividend policy is not the same as they expected.

[7] https://fundresearch.fidelity.com/mutual-funds/fees-and-prices/316138106.

2.3 Hypotheses

2.3.1 Does Multi-Dividend Policy Hurt Investors?

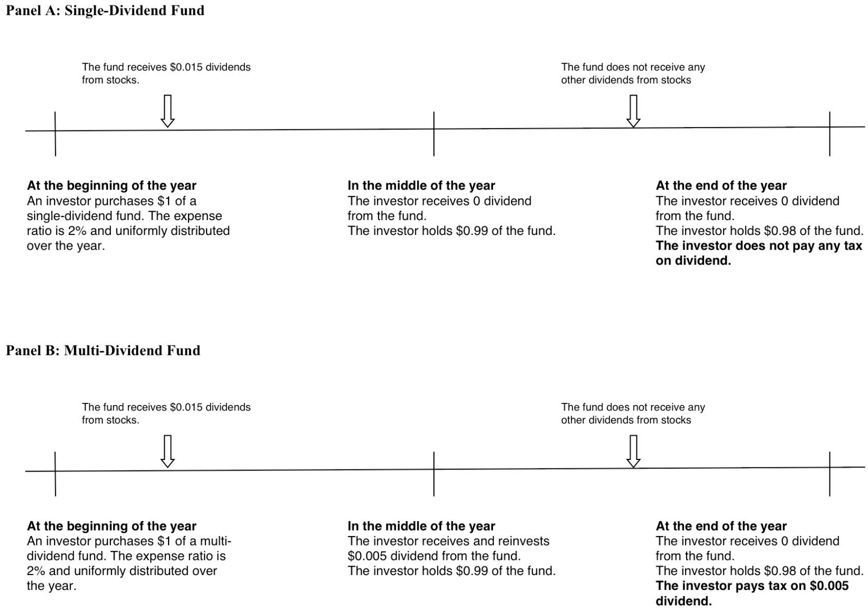
Dividend distribution is not cost-free. When a mutual fund passes dividends to its shareholders, it incurs transaction fees (e.g., transference costs). Some of the costs might be fixed for each distribution. Therefore, multi-dividend funds might need to pay more cost than single-dividend funds. These transaction fees are ultimately paid by shareholders. As a consequence, one could expect multi-dividend funds on average charge higher transaction fees than single-dividend funds. The extra charged fees may erode fund performance.

Hypothesis 1 *Multi-dividend funds underperform single-dividend funds, other things equal.*

The costs that multi-dividend policy brings is not always explicit. One important dimension of hidden costs from multi-dividend policy is tax. Figure 2.1 illustrates the potential cost of investing in a multi-dividend fund resulting from taxes. We assume two identical funds with different dividend policies (semiannual and annual). The expected return is zero and the reinvestment rate is 100%. The funds in both panels have the same expense ratio, 2%, and receive the same amount of dividends, 1.5%. We assume that the expense is uniformly charged during the year. Both funds receives \$0.015 dividends per each dollar from stocks in the first half year and \$0 in the second half year. At the beginning of year, an individual invests \$1 in each fund in panel A and panel B. In the middle of the year, the investor receives 0 dividend from fund A and \$0.005 from fund B. The investor holds \$0.99 of each fund. At the end of the year, the investor receives 0 dividend from each fund and holds \$0.98 of each fund. The investor pays tax on \$0.005 dividends received from fund B. As a results, the before-tax payoffs are the same for both funds in panel A and B. Yet, the fund A dominates the fund B since the investor pays taxes on the dividends that she receives.

The tax cost of multi-dividend policy results from the mismatch between the period of personal income tax and that of dividend distribution, i.e., personal income tax is calculated on a yearly basis while the dividends are distributed on a shorter basis. Therefore, a multi-dividend fund could avoid potential tax cost caused by dividend policy if the outcomes of dividend payout (e.g., pay

Figure 2.1: Time line of taxes paid under different assumptions about the dividend policy of a mutual fund



or not pay) in each period are the same as the outcomes when the fund uses the single-dividend policy. It implies that the dividend payments are stable over the time, i.e., a fund keeps paying dividends or not over the time.

There are two ways for multi-dividend funds to stabilize dividend payments. The first way is to stabilize the dividends that a fund receives from the stocks, given the fees charged by funds are relatively fixed. It is not difficult to predict how many and when a stock is going to pay dividends from its characteristics and past dividend history (Fama and French, 2001; Hartzmark and Solomon, 2013). The funds, therefore, could make the corresponding adjustments on their holdings and generate the amount of dividends they wish. Harris et al. (2012) provide evidence that some mutual funds change their holdings before dividend payment (ex-day). The dividend stability also stems from the pressure of investors. In section 2.3.3, we argue that individuals invest in multi-dividend funds for the purpose of obtaining constant and stable dividend flows. If a multi-dividend fund targets these investors, it should keep net dividend stable in terms of quantity in each dividend period or the investors would flee away for it. However, single-dividend funds are not attractive for these investors, and consequently, do not have any incentive to stabilize the dividend because it is costly.

A mutual fund can also keep dividend payment stable through setting its expense ratio. For example, a fund could set a low expense ratio if it hopes to pay dividends to investors. As such, a multi-dividend fund could charge low fees, including waive or reimburse fees, to generate more available dividends passing through to its shareholders if the fund could expect the future dividends would be stable and low. This operation actually benefits the investors. Meanwhile, a fund has no incentive to charge high fees while it uses multi-dividend policy.

Hypothesis 2A *Multi-dividend funds have more stable dividend ratios than single-dividend funds, other things being equal.*

Hypothesis 2B *The dividend frequency is negatively related to expense ratio, other things being equal.*

2.3.2 Conflict between Existing and New Investors

Figure 2.2 illustrates that existing and new investors might have different preferences on dividend policy. Assume that two identical funds, X and Y, have different dividend policies (annual and semiannual). We assume that the expected return is 0, the total net asset (TNA) is \$100, and the reinvestment rate is 100%. Both funds only have one shareholder A. Funds receive \$10 dollars as dividend at the beginning of the year and no other dividends in the rest of year. X would pay out the dividends at the end of the year, and Y would pay out in the middle of the year. A new investor B invests \$20 dollars into both funds (the relative flow is 20%) after Y pays out the dividends. Investors A and B would have the same payoffs from fund A and B, other things being equal. However, investor B has to pay the taxes for the dividends received from fund X. Table 2.1 shows the payoffs of both investors from X and Y from two funds. For existing investor A, fund X, which uses single-dividend policy, is better than multi-dividend fund Y. On the other hand, new investor B would find multi-dividend fund Y is more favorable. It shows that the dividend policy most attractive to new investors may be costly for existing investors. Hence, compared with single-dividend policy, multi-dividend policy might attract new investors and, consequently, increase mutual fund's inflows. Yet, multi-dividend policy also brings potential costs to existing investors. A mutual fund needs to trade off interest between existing and new investors to maximize its size.

Mutual funds pay distributions (e.g., dividends) following the equal allocation rule, i.e., each share would receive the same amount of dividends. Investors

Figure 2.2: The Impact of Dividend Policy on After-tax Returns

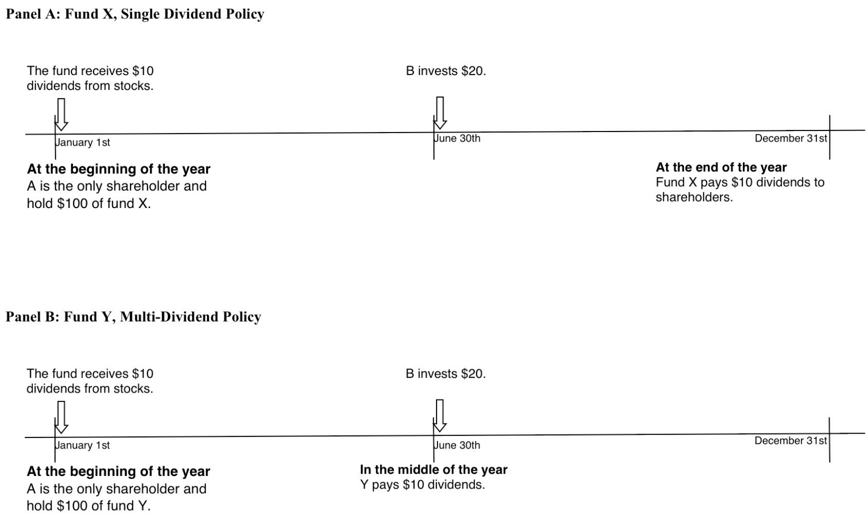


Table 2.1: Payout of Exiting and New Investors

Investor	Fund	Dividend	Tax	Returns
A	X	\$8.33	\$2.75	-2.75%
	Y	\$10	\$3.3	-3.3%
B	X	\$1.67	\$0.55	-2.75%
	Y	\$0	\$0	0

as of a distribution date would share dividends according to the proportion of shares they hold in the whole portfolio, regardless of their purchase date or whether their shares appreciate or not. That is to say, a new investor would share the undistributed dividends accumulated from the nearest dividend distribution to purchase time with existing investors. It follows that a new investor would receive more dividends in the current year and pay unnecessary income taxes. The investor would try to avoid funds with a high overhang of undistributed dividends. This argument is in spirit to that of Bergstresser and Poterba (2002). They find a fund with heavy-taxed returns (e.g., undistributed dividends), receives lower inflows than funds offering similar pretax returns but lower tax burdens. Johnson and Poterba (2010) show that retail investors time their purchase of mutual fund to avoid tax acceleration with distributions. If more frequent dividend distributions could lower the tax burdens of new investors, mutual funds would have incentive to pass through dividends to relieve dividend overhang as soon as possible, such as using multi-dividend policy.

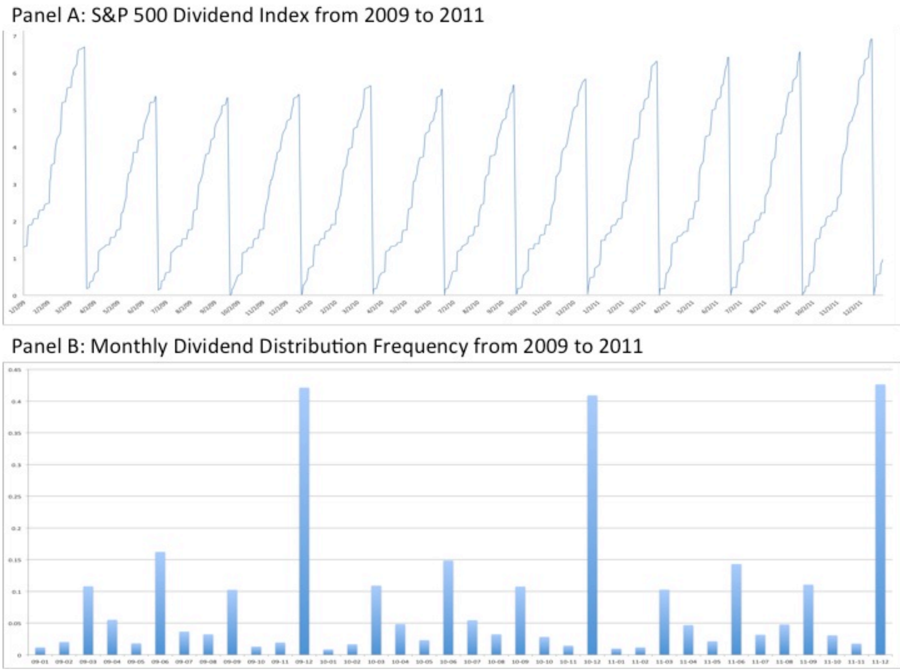
The equal allocation rule also causes unrealized capital gain overhang (UCGO) (Barclay et al., 1998; Bergstresser and Poterba, 2002). Dividend overhang (DO) in this chapter is different from UCGO in two aspects. First, DO affects new investors by transferring capital gain taxes in the future to dividend income taxes at the current year, or vice versa for existing investors. Because tax rates for capital gains and dividend income are different, DO changes tax burdens not only in time but also in quantity. Yet UCGO only retimes the capital gains tax liability of investors, if any, but not tax quantity (Bergstresser et al., 2003). As such, some short-term investors may not be affected by UCGO but by DO. Second, DO can have a larger impact on low-income investors. If an investor' income tax rate is 10% or 15%, his or her capital gain rate is 0%. As a result, the investor needs to pay the taxes on dividend income but not on long-term capital gains.

The dividend overhang problem would exist in a fund unless a fund pays out undistributed dividends every day. Figure 2.3 intuitively shows when a mutual fund receives dividends and when it distributes them. Panel A shows how the S&P 500 dividend index evolves from January, 2009 to December, 2011[8]. Panel B illustrates the frequency of dividend distribution each month during the same period. It is easy to observe that a mutual fund typically pays out dividends at the end of year/quarter, but it receives dividends all the time.

Assume a fund receives dS from the underlying assets it holds, where d is the dividends received for each dollar in the portfolio and S is the total net assets of the portfolio held by existing investors in dollar. We assume the expense ratio is 0. Thus, the fund would pay dS of dividends. Just before the distribution, a new investor buys s dollars of this fund. As a consequence, the new investor would receive $\frac{dS}{S+s}$ per dollar as dividends. At the end of the year, the investor needs to pay $\frac{dS}{S+s}t_d$ as income tax for each dollar invested, where t_d is the income tax rate. If the dividend ratio $\frac{dS}{S+s}$ is, for example, 1% and the income tax rate t_d is 33%, the new investor loses 0.33% of the principal. This number increases with the dividend ratio. So the potential tax cost for the new investors is positively related to dividend ratio. As a result, new investors might avoid buying funds with high dividend ratios. Consequently, those funds

[8]The S&P 500 dividend index measures the total dividends paid in the underlying index since the previous rebalancing date. The index resets to zero on a quarterly basis. The data is downloadable at S & P index website.

Figure 2.3: S&P 500 Dividend Index and Dividend Distributions
Panel A of this figure illustrates how S&P 500 dividend index evolves from 2009 to 2011. Panel B shows the monthly frequency of dividend distributions from 2009 to 2011. S&P 500 dividend index measures the total dividends paid in the underlying index since the previous rebalancing date. The index resets to zero on a quarterly basis. The monthly dividend distribution frequency is calculated as the times of dividend distributions during one month over the total times during the year.



have incentives to pay dividends more frequently (e.g., multi-dividend policy) to attract new investors.

Because of the equal allocation rule, existing investors receive the same amount of dividends as new investors, $\frac{dS}{S+s}$ per dollar, and pay the corresponding income tax. In another scenario without new investors, existing investors would receive d per dollar. The difference of the dividends received by existing investors in two scenarios is $d - \frac{dS}{S+s} = \frac{ds}{S+s}$. Existing investors always receive less dividends in the scenario with the existence of new investors. Therefore, existing investors would prefer a fund with larger dividend ratios to use single-dividend policy so that more new investors could share the dividends. They would prefer the single-dividend funds since they could receive more dividends from multi-dividend funds than from single-dividend funds, other things being equal.

Yet, it is worthy to note that existing investors are less affected by dividend policy than new investors. With the aforementioned notations, the difference of taxes an existing investor would pay is $\frac{ds}{S+s}t_d$ per dollar in the current year, whereas the difference for a new investor between two scenarios is $\frac{dS}{S+s}t_d$. Therefore, a new investor incurs more unnecessary costs for choosing wrong dividend policy than existing investors in the current year ($\frac{ds}{S+s} < \frac{dS}{S+s}$) when the relative flows are not too huge ($\frac{s}{S} < 1$)[9]. Moreover, existing investors still need to pay capital gain taxes on the dividends shared by new investors. They swap income taxes in the current year for capital gain taxes in the future in the scenario with new investors. When they sell the funds in the future, they have to pay $\frac{ds}{S+s}t_{cp}$ per dollar as capital gain tax. Therefore, without considering money's time value, the difference of taxes for existing investors between two scenarios turns to $dt_d - \frac{dS}{S+s}t_d - \frac{ds}{S+s}t_{cp}$, or $\frac{ds}{S+s}(t_d - t_{cp})$, where t_{cp} is the capital gain tax rate and $t_d > t_{cp}$. This result suggests that dividend policy has much smaller impact on existing investors than on new investors, and therefore, exiting investors are less sensitive to the dividend policy than new investors. Funds might benefit from multi-dividend policy by attracting new investors and not losing too many existing investors.

Hypothesis 3 *Mutli-dividend policy is positively associated with dividend ratio ($\frac{dS}{S+s}$), controlling for other characteristics.*

Some investors in the market may seek regular income streams. They might buy fixed-income assets, i.e., money or bond market funds, bank accounts, and short-term paper, since they distribute dividends more frequently and stably. However, it does not favor mutual fund management companies' interest. The expenses charged by equity funds are much higher than that by fixed-income assets, such as bond funds. Therefore, fund management companies might encourage money flows from fixed-income assets to equity funds.

One way to attract money from fixed-income assets is through divided policy. As Fidelity Investments posts in its website, "with today's rates already very low, bond market return dynamics may look different moving forward, and these changes may help to make dividends (of equity funds) look attractive."[10] Some investors even might equal multi-dividend equity funds to fixed-income assets without considering their risks are different[11].

Equity funds face two competitions for intriguing investors from fixed-income assets. The first competition is between equity funds and fixed-income assets.

---

[9] In my sample, more than 96% observations have relative flows smaller than 1.

[10] https://www.fidelity.com/learning-center/trading/all-about-dividends/new-era-for-dividends

[11] For example, CNN reports that some investors might mistake multi-dividend funds as fixed-income portfolios. See http://money.cnn.com/2011/11/03/pf/expert/bond_funds/index.htm

In equilibrium, equity funds might increase dividend frequency when fixed-income assets are more attractive, i.e., yields on fixed-income assets are higher, and vice versa. The second competition is within the equity funds. When yields on fixed-income assets are lower, money flows from fixed-income assets to equity funds. As such, an equity fund needs to compete with other equity funds by increasing the dividend frequency. However, the domination of these two competitions is still unclear.

We use risk-free interest rate to proxy for the return of fixed-income assets. Admittedly, the return on fixed-income assets is plausibly associated with the corporate dividend payment and, subsequently, affects the dividends received by mutual funds. In this chapter, we do not consider it is important because (1) the impact of corporate dividend payment is the same for multi-dividend and single-dividend funds and (2) the impact of interest rate on corporate dividend payment is indirect whereas that on fixed income asset returns is direct.

Hypothesis 4 *Mutual funds are more likely to pay dividends more frequently when the return of fixed-income assets is higher, controlling for other fund characteristics.*

2.3.3 Conflict within Existing Investors

Existing investors of mutual funds have different interests. Mutual funds trade off interest within existing investors in order to maximize their size and, therefore, revenue (Christoffersen et al., 2005). Similarly, there are reasons to believe that existing investors have different preferences in dividend policy. Previous studies (Becker et al., 2011; Graham and Kumar, 2006; Scholz, 1992; Thaler and Shefrin, 1981) well document the existence of dividend clientele. These authors propose that some investors might prefer high dividend stocks for consumption or tax purposes.

Previous empirical studies based on dividends do not distinguish between consumption and tax rationale. The investors who have more pronounced needs to finance their consumption typically have lower tax rates. Mutual fund dividend policy provides a unique opportunity to test the purpose of investment from dividend clientele. Because of the pass-through regulation, the sum of the dividends for a year would be similar for most funds with small relative flows (see discussion in the previous section). Therefore, if the demand for dividends stems from that some investors have a relative low tax rates, they would not show preference to multi-dividend funds because their tax burdens would be almost the same. Contrary to tax rationale, under the assumption that an investor uses dividends to finance his or her consumption, he or she would prefer regular income streams, as in multi-dividend funds, because of self-control (Thaler and Shefrin, 1981). The following argument in this section is under the consumption rationale.

If the purpose of investing in a mutual fund is to gain regular income streams, investors can realize this purpose via two ways related to equity funds: holding multi-dividend funds and partially selling single-dividend funds. Under the assumptions of Miller and Modigliani (1961), the dividend policy is irrelevant to investors' choices in the absence of transaction costs since they could generate regular dividend streams by themselves. Investors could always cancel out a firm's dividend policy by realizing "homemade dividends", i.e., partly selling shares. However, redeeming shares could be expensive. One type of cost is related to the information cost. To redeem shares, they probably want to forecast the future returns of the funds and choose the time to sell the funds. As such, investors need to actively or passively collect and analyze information. They might even regret after they sell the funds if they made the wrong decisions. Another type of cost is related to transaction cost. Investors need to pay, if any, front/back-end loads, brokerage fees, when they sell and reinvest the funds. There is also the opportunity cost of time spent trading shares. Following Huang et al. (2007), I term those costs participation costs. As such, investors who are seeking for regular income streams might avoid investing in funds with high participation costs if they use a single-dividend policy. In equilibrium, a multi-dividend fund is more likely to have participation costs than a single-dividend fund if they both target dividend clientele. Admittedly, some single-dividend funds may not target dividend clientele. Therefore, their investors are more likely to pose obtaining better fund performance as the primary goal of investing funds and, consequently, be performance-sensitive. They buy and sell funds more frequently. As such, those investors would avoid single-dividend funds with high participation costs. As a result, a multi-dividend fund has higher participation costs than a single-dividend fund.

Hypothesis 5 *Dividend frequency is positively associated with participation costs.*

2.4 Data

I obtain data from the CRSP Survivor-Bias Free Mutual Fund Database spanning from 2000 to 2011. The original sample contains all open-end mutual funds that are active from 2000 to 2011. From the initial sample, I retain domestic equity mutual funds defined by the lipper objective codes.[12] I also exclude index and institutional funds identified by CRSP identifiers, i.e., *index_fund_flag* and *inst_fund*. I identify funds with multiple share classes by *crsp_cl_grp* provided by CRSP and compute the fund characteristics as the asset-weighted means of class characteristics. In some rare cases, the dividend payment varies across classes even in the same fund. It stems from that classes in the same portfolios have different fee structures. For example, assume one portfolio has two classes with expense ratios 1.55% and 1.50%, respectively. The dividend ratio before fee is 1.52%. As a consequence, one class in this portfolio pays dividends whereas the other class does not. I define a fund pays a dividend payment if any class in its portfolio pays. Following Elton et al. (2011) and Evans (2010), I drop the smallest funds, i.e., total net assets below $15 million, and the young funds, i.e., the age is less than 36 months. This leaves a sample of 3,257 distinguish funds and 18,574 fund-year observations.

I am interested in examining how a mutual fund determines the frequency of dividend distributions that the fund intends to pay out. As such, I estimate dividend policy as the realized frequency of dividend distributions during a calendar year plus one if a fund does not pay dividends in December and zero otherwise. There are two considerations for the definition. First, we use realized dividend frequency, rather than scheduled dividend policy, to calculate dividend policy. Empirically, realized dividend frequency is typically consistent with intended dividend frequency. In some rare cases, a fund does not pay dividends to its shareholders if the fund does not collect enough dividends to offset fees. However, funds can easily avoid this situation. A mutual fund can simply buy stocks before the dividend distribution date to collect enough dividends (Harris et al., 2012). As such, realized dividend frequency is representative of mutual fund intention. Another reason is that the real dividend policy that a fund intends to use is difficult to observe. Not all funds report scheduled dividend frequency in the prospectus and funds might change their intentions over time. Second, a mutual fund can choose fiscal or calendar year as its taxable year.[13] However, excise tax rule suggests that all mutual funds need to distribute, if any, dividends at the end of December. Therefore, we assume that all mutual funds intend to apply this rule, i.e., they all plan to pay dividends in December, even if it ends up that they do not pay any dividend in December. Empirical results support this assumption. In my sample, more than 98% mutual funds that have one dividend distribution choose to pay it in December. I use the realized frequency of dividend distributions and the

[12] We consider the fund is a domestic equity fund when the fund is classified as one of the following categories by *lipper_class*: LCVE, MLVE, EI, EIEI, LCCE, MLCE, LCGE, MLGE, MCVE, MCCE, MCGE, SCVE, SCCE, and SCGE.

[13] http://www.irs.gov/Businesses/Small-Businesses-%26-Self-Employed/Tax-Years

expected dividend frequency, i.e., the maximal dividend frequency over the past three years. The results are robust.

Table 2.2 provides the summary statistics for our sample. All definitions of the variables of fund characteristics are similar to previous literature. My main variables are dividend frequency and dividend ratio. Some mutual funds pay dividends[14] several times in the same month or in the same day, typically when dividends belong to different types (e.g., income dividend and qualified income dividend). Therefore, I consider dividend distributions in the same month as one time. I calculate dividend frequency as the numbers of months when a mutual fund pays dividends (from 0 to 12). I define multi-dividend dummy equals to 1 if dividend frequency is larger than one time, and zero otherwise. The yearly dividend ratio is calculated as the sum of dividend ratios, defined as the distribution amount over reinvestment price, for each dividend distribution throughout the year.

In panel A of table 2.2, I report the summary statistics for all funds in the sample. Among those funds that at least pay dividend once in the year, nearly 30% are multi-dividend funds. Funds, on average, pay dividends 1.7 times per year. Panel B provides the mean and the standard deviation of variables of interest for funds paying dividends one, two, three or four times and larger than four times: 70.4% fund-years use single-dividend policy and pay dividend once per year. 14.93% and 12.64% fund-years pay dividends twice and three or four times per year, respectively; and 2% fund-years pay dividends more than four times. In panel B, apart from the observations with dividend frequency equal to 1, dividend ratio, load, and risk-free interest rate increase with dividend frequency whereas the past 12-month raw return is negatively associated to dividend frequency, suggestive of our hypotheses. The univariate relationship between variables of interest and dividend policy fits our hypotheses across all dividend frequencies. In the robustness test (unreported), I drop the observations that there is no dividend payment during the year to avoid suspicious missing observations. The results are mainly the same.

Since mutual fund segments have their own objectives, they might have the corresponding dividend policies across segments. To offer a broader view of the database, Table 2.3 reports the summary statistics relating to fund segments. Panel A reports the summary statistics of dividend frequency and dividend ratio in each segment. The first observation is that mutual fund dividend policy varies across mutual fund segments. Consistent with my expectation, dividend frequency is highly associated with dividend ratio. Income, large-cap funds, which have large dividend ratios, pay dividend distributions more frequently. Growth and small-cap funds are more likely to use single-dividend

[14] I identify dividend distributions as mutual fund dividends if the first letter of *dis_type* in CRSP is "D".

Table 2.2: Summary Statistics
This table presents the summary statistics for our database sample. Panel A provides the summary statistics for variables of interests. Panel B reports the summary statistics across the dividend frequency. DivRatio is the sum of the dividend ratio, defined as distribution amount over reinvestment price, for each dividend distribution during the whole year. LnDivRatio is the natural logarithm of the dividend ratio plus 1. DivFreq is the number of months when a fund pays dividends. MultiDiv is the dummy variable that equals 1 if DivFreq is larger than 1, and 0 if DivFreq is equal to 1. LnSize is the nature logarithm of total net asset under fund management. LnAge is the nature logarithm of age in months. FrontLoads is the front load. BackLoad is the back load for holding 48 months. ExpRatio is the expense ratio defined as total operating expenses divided by the year-end total net assets. TurnRatio is the turnover ratio. FundRet is the raw return for the past 12 months (buy and hold). Size is the total net asset under fund management. RF is the risk-free interest rates, defined as the one-month Treasury bill rate at the end of the year.

PANEL A: Summary Statistics for variables of interests

	Mean	SD	Median	1st perc.	99th perc.	N
DivRatio (%)	0.498	1.084	0.026	0.000	3.666	18574
LnDivRatio (%)	0.492	1.022	0.026	0.000	3.601	18574
DivFreq (times)	1.690	1.560	1	1	12	18574
MultiDiv	0.296	0.457	0	0	1	18574
LnSize ($ million)	5.856	1.613	5.752	2.918	9.997	18090
LnAge (month)	4.946	0.658	4.875	3.871	6.758	18574
FrontLoad (%)	1.766	2.118	0.000	0.000	5.750	18217
BackLoad	0.116	0.276	0.000	0.000	1.272	18217
ExpRatio (%)	1.321	0.430	1.315	0.203	2.413	16213
TurnRatio (%)	85.049	87.257	64	3	401	16097
Fundret (%)	4.382	20.849	7.427	-45.024	46.689	14826

PANEL B: Summary Statistics across Dividend Frequency

	DivFreq				All
	1	2	3 or 4	≥5	
N	13073	2773	2348	380	18574
DivRatio	0.250	0.985	1.096	1.801	0.498
(% per year)	(0.612)	(1.563)	(2.78)	(2.822)	(1.084)
LnSize	5.727	6.046	6.364	5.741	5.856
($ million)	(1.565)	(1.631)	(1.762)	(1.354)	(1.613)
LnAge	4.904	4.982	5.139	4.909	4.946
	(0.633)	(0.646)	(0.761)	(0.668)	(0.658)
FrontLoad	1.839	0.902	2.189	2.829	1.766
(%)	(2.102)	(1.781)	(2.276)	(2.067)	(2.118)
BackLoad	0.131	0.0399	0.112	0.175	0.116
(%)	(0.292)	(0.158)	(0.263)	(0.311)	(0.276)
ExpRatio	1.398	1.056	1.220	1.272	1.321
(% per year)	(0.417)	(0.423)	(0.370)	(0.341)	(0.430)
FundRet	4.307	5.289	4.200	2.475	106.550
(% per year)	(21.895)	(18.049)	(18.370)	(17.929)	(19.86846)
Rf	0.158	0.106	0.171	0.179	0.152
(% per year)	(0.151)	(0.144)	(0.156)	(0.150)	(0.152)

Table 2.3: Summary Statistics for Segments
Panel A summarizes dividend frequency across segments in the sample. Panel B provides the proportion of each segment in the sample over years. EIEI is equity income funds. LCVE is large-cap value funds. LCCE is large-cap core funds. MLVE is multi-cap value funds. MLCE is multi-cap core funds. MCVE is mid-cap value funds. MCCE is mid-cap core funds. SCVE is small-cap value funds. LCGE is large-cap growth funds. SCCE is small-cap core funds. MLGE is multi-cap growth funds. MCGE is mid-cap growth funds. SCGE is small-cap growth funds. DivFreq is the dividend frequency defined as frequency of dividend distribution during one year. DivRatio is the dividend ratio defined as sum of dividend distribution amount over the reinvestment price reported by funds during one year.

PANEL A: Dividend Frequency across Segments

Lipper Code	Class Name	DivFreq				DivRatio (%)	Obs.
		Mean	10%	Median	90%		
EIEI	Equity Income Funds	5.021	1	4	12	1.884	746
LCVE	Large-Cap Value Funds	2.486	1	2	4	1.066	1226
LCCE	Large-Cap Core Funds	2.132	1	1	4	0.741	2702
MLVE	Multi-Cap Value Funds	2.119	1	1	4	0.840	1201
MLCE	Multi-Cap Core Funds	1.579	1	1	3	0.656	2118
MCVE	Mid-Cap Value Funds	1.543	1	1	3	0.475	678
MCCE	Mid-Cap Core Funds	1.449	1	1	2	0.384	980
SCVE	Small-Cap Value Funds	1.337	1	1	2	0.391	861
SCCE	Small-Cap Core Funds	1.270	1	1	2	0.271	1713
LCGE	Large-Cap Growth Funds	1.253	1	1	2	0.153	2163
MLGE	Multi-Cap Growth Funds	1.106	1	1	2	0.010	1311
MCGE	Mid-Cap Growth Funds	1.051	1	1	1	0.037	1432
SCGE	Small-Cap Growth Funds	1.029	1	1	1	0.040	1443

Panel B: Percentage of Mutual Fund Types over Years (%)

Lipper Code	Year											
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
EIEI	5.274%	4.991%	4.208%	3.743%	3.331%	3.421%	3.141%	3.641%	3.979%	4.350%	4.180%	4.268%
LCCE	10.348%	13.047%	16.172%	18.182%	17.071%	15.359%	14.071%	12.549%	12.268%	13.712%	15.049%	15.619%
LCGE	10.846%	14.711%	12.954%	10.466%	10.132%	9.926%	10.537%	10.013%	11.207%	13.191%	12.912%	12.054%
LCVE	9.652%	6.042%	6.271%	5.882%	5.413%	5.701%	6.283%	6.307%	7.692%	6.761%	6.456%	7.176%
MCCE	3.184%	3.590%	3.960%	4.966%	6.107%	6.170%	5.497%	5.657%	5.172%	5.768%	6.224%	5.113%
MCGE	6.667%	7.968%	8.333%	8.403%	7.981%	8.317%	8.246%	8.323%	7.958%	7.612%	6.921%	6.567%
MCVE	5.672%	4.116%	3.795%	3.361%	3.331%	3.957%	4.450%	4.226%	4.178%	3.168%	2.183%	3.143%
MLCE	8.060%	6.743%	7.838%	9.626%	11.520%	12.072%	12.042%	13.199%	11.804%	12.199%	13.423%	13.180%
MLGE	11.841%	8.319%	5.776%	5.653%	6.801%	6.908%	6.283%	7.412%	7.095%	6.478%	6.595%	7.317%
MLVE	11.045%	11.384%	9.158%	8.327%	6.870%	6.707%	7.003%	5.722%	4.973%	4.303%	4.134%	4.268%
SCCE	4.279%	5.079%	6.353%	8.480%	9.160%	9.457%	9.817%	10.533%	10.809%	10.260%	11.426%	9.991%
SCGE	5.771%	7.706%	8.168%	7.945%	8.536%	8.250%	8.312%	8.062%	8.157%	8.038%	7.106%	7.083%
SCVE	7.363%	6.305%	7.013%	4.966%	3.747%	3.756%	4.319%	4.356%	4.708%	4.161%	3.391%	4.221%

policy. However, the relationship between dividend frequency and dividend ratio is not perfect. For example, although the funds in segment "MLVE" have larger dividend ratios than the ones in segment "LCCE", the average of dividend ratio in the former segment is smaller than that in the latter segment. The second observation is that the dividend policy varies across mutual funds within the same segment. Even in the segment with the lowest dividend ratio, there are some funds pay dividends more frequently than the mandated time. Panel B provides the percentage of mutual fund segments changes over time. The largest five segments in terms of percentage in 2011 are "LCCE", "MLCE", "LCGE", "SCCE" and "LMGE". They represent segments with different dividend policies.

2.5 Empirical Strategy and Results

2.5.1 Dividend Distribution Frequency as Strategic Choice

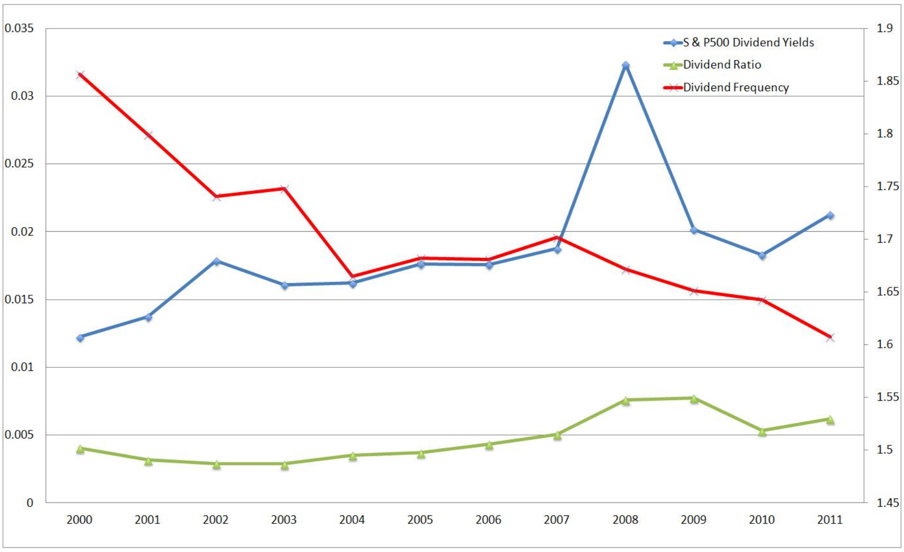
I begin my empirical test by examining whether mutual funds consider dividend distribution frequency as a strategic choice. Under this hypothesis, we should expect that a mutual fund's dividend distribution frequency would depend on its strategy rather than market characteristics (e.g., market dividend yields). The first prediction is that mutual funds' dividend policy is persistent from year to year. If a fund chooses its dividend policy randomly, the frequency of dividend distributions would vary over years. Many mutual funds report their dividend distribution schedules in their prospectuses. This promise, if any, to a large extent, ensures the persistence of dividend policy. To formally examine the persistence of dividend policy, following the methodology in Carhart (1997) and Berk and Tonks (2007), I count the numbers of funds that keep (change) their dividend policy from year t to year t+1, as well as from year t to year t+2. Table 2.4 reports the results. The results show that 91% of mutual funds have persistent dividend policy in the next year and 90% of funds' dividend policy remains persistent in the year t+2.

Table 2.4: Persistence of Dividend Frequency
This table reports the number of the funds using different dividend policies. Multi-dividend implies a fund uses multi-dividend policy. Single-dividend implies a fund uses single-dividend policy.

Year t	Year t+1		Year t+2	
	Multi-dividend	Single-dividend	Multi-dividend	Single-dividend
Multi-dividend	3904	550	2922	549
Single-dividend	566	10297	574	8364

Another prediction is that the frequency of dividend distributions is, to some extent, independent from the market characteristics. Figure 2.4 illustrates that the market average frequency of dividend distribution is irrelevant to the stock market dividend payout. It shows how the average mutual fund dividend frequency, the dividend ratio, and the S&P 500 dividend yield evolve over time. S&P 500 dividend yield (blue line) increases from 2000 to 2011 with a peak in 2008. The dividend ratio (bottom line) has a shape similar to S&P 500 dividend yield, but smoother. If a mutual fund does not manipulate the dividend distribution frequency, the mutual fund average frequency and the market dividend yield are supposed to follow the same distribution and increase. However, to the contrary, the average frequency of dividend distribution (top line) gradually decreases from 2000 to 2011. However, this result might be driven by the decrease of the percentage of segments with high dividend distribution frequencies in the market.

Figure 2.4: Dividend Policy over Time
The figure illustrates how average mutual fund dividend frequency, dividend ratio, and S&P 500 dividend yield evolve over time. Middle line is the dividend payout of stocks in S&P 500 index. Bottom line is average fund dividend ratio, defined as the cross-sectional mean of dividend ratio. Top line is the average fund dividend frequency, defined as the cross-sectional mean of dividend frequency.



To further examine this prediction, I consider the only possible scenario that a mutual fund does not follow the schedule to pay dividends: when a mutual fund does not collect enough net dividend, defined as dividends received subtract fees, it would pay nothing to its shareholders in the scheduled dates. Therefore, the mutual fund has a lower dividend frequency than scheduled. It is supposed to be common in the market if a mutual fund does not have any dividend preference. A typical expense ratio for a mutual fund is 1.5%, and the S&P 500 dividend yields range from 1.1% to 3% in the period from 2000 to 2011.[15] In my sample, 27.66% of the fund-year's expense ratio is higher than the S&P 500 dividend yield. Therefore, the percentage of the fund-year, which pays dividend distributions less frequently than the scheduled, proxied by the maximum distribution frequency in the past three years (from t-2 to t), would be around this number. However, this number is only 6.85%. This large

[15]Source: Shiller (2006), www.irrationalexuberance.com

difference implies that mutual funds do not choose dividend policy randomly. They deliberately choose dividend policy and keep it persistent.

2.5.2 Dividend Policy and Fund Characteristics

In this section, we examine the hypotheses in section 2.2, i.e., how a mutual fund determines its dividend policy. I model multi-dividend policy as a function of fund characteristics, which are associated with how mutual funds resolve the conflicts of interest between existing and new investors as well as within existing investors. Table 2.5 reports the results of our multivariate regressions. Columns (1), (3), (5) and (7) present the results for the logit regressions of the dummy for multi-dividend policy on fund characteristics. The dependent variable is *Multidiv*, which equals 1 if a fund pays dividends more than one time during one year and 0 if a fund only pays dividend once. The first, the second, and the third rows for every variable correspond the coefficient, the marginal effect, and the test statistics (Z-statistics). Columns (2), (4), (6) and (8) report the coefficients and t-statistics for the pooled OLS regressions of the dividend frequency in the current year on fund characteristics. Standard errors in all columns are clustered by fund and year. Regressions in the first four columns include year and segment fixed effects. The other two regressions include segment fixed effects.

We start with examining whether multi-dividend policy potentially hurts investors. The coefficients of *FundRet* is negative in columns (1), (2), (7) and (8), although not significant in columns (2) and (8), implying multi-dividend policy is associated with worse raw returns. Meanwhile, the coefficients on *FundRetRA* are insignificant in columns (3) and (4) suggesting risk-adjusted returns are not different across dividend policy. The results provide very limited evidence supporting Hypothesis 1. The transaction costs caused by dividend policy might be marginal to the performance. One possible explanation of the association between raw returns and dividend policy is that multi-dividend funds may continuously keep a large proportion of low-return assets (e.g., cash) in their portfolios and therefore get a low raw returns. The dividends can only be paid in the form of cash. Funds need to keep larger proportion of cash in the portfolio before each distribution. As such, the portfolio in multi-dividend funds on average are more liquid than that in single-dividend funds.

To investigate whether multi-dividend policy leads to a more stable dividend income as suggested by Hypothesis 3, I construct a new variable *Stability*

Table 2.5: Fund Characteristics and Dividend Policy
This table reports the regressions of dividend policy on fund characteristics. Columns (1), (3), (5) and (7) report the coefficients, marginal effects, and associated z-values from logit regressions for the probability of a fund using the multi-dividend policy. The first, second, and third rows of each variable reports coefficients, marginal effects and z-statistics, respectively. The dependent variables are dummy variables equal to 1 if a fund uses multi-dividend policy. Columns (2), (4), (6) and (8) report the coefficients and t-statistics from OLS regressions of dividend frequency on fund characteristics. Stability is the natural logarithm of the absolute value of the ratio of the difference between dividend ratio at t and dividend ratio at t-1 to dividend ratio at t-1 plus 1, defined in equation 2.1. FundRisk is the standard deviation of raw returns during one year. FundRetRA is the average monthly risk-adjusted alpha during the year. All other variables are defined in Table 2.2. Year FE and Segment FE shows whether we include year and segment fixed effects. Standard errors are clustered by fund and year. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.

	Dividend Policy							
	LOGIT	OLS	LOGIT	OLS	LOGIT	OLS	LOGIT	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{LnDivratio}_{i,t}$	35.370***	37.959***	40.197**	33.384***	27.896***	34.256***	36.036***	38.005***
	8.489		9.775		6.706		8.649	
	(3.33)	(4.72)	(2.45)	(2.88)	(2.9)	(4.75)	(3.44)	(4.70)
$\text{Stability}_{i,t}$	0.452***	0.473***	0.553***	0.460***	0.446***	0.470***	0.443***	0.464***
	0.109		0.134		0.107		0.106	
	(4.98)	(7.39)	(5.21)	(6.67)	(4.99)	(7.35)	(4.87)	(7.26)
$\text{ExpRatio}_{i,t}$							-0.280	0.049
							-0.067	
							(-1.08)	(0.31)
$\text{No12-1B}_{i,t}$	-0.424	-0.017	-0.460	-0.122	-0.470*	-0.049		
	-0.102		-0.112		-0.113			
	(-1.59)	(-0.09)	(-1.49)	(-0.56)	(-1.83)	(-0.27)		
$\text{PartCost}_{i,t}$	0.254**	0.503***	0.277**	0.472***	0.183*	0.450***		
	0.061		0.067		0.044			
	(2.15)	(4.05)	(2.35)	(3.92)	(1.69)	(3.94)		
$\text{FundRet}_{i,t}$	-0.010**	-0.004			-8.993*	-7.670*	-0.011**	-7.670
	-0.003				-2.162		-0.003	
	(-2.00)	(-1.28)			(-1.90)	(-1.65)	(-2.08)	(-1.43)
$\text{FundRetRA}_{i,t}$			-0.010	-0.002				
			-0.002					
			(-1.17)	(-0.53)				
RF_t					-0.006**	-0.003		
					-0.002			
					(-2.43)	(-1.03)		
$\text{FrontLoad}_{i,t}$							0.086**	0.148***
							0.021	
							(2.41)	(4.00)
$\text{BackLoad}_{i,t}$	-0.040	0.365	-0.103	0.357	0.192	0.577**	0.153	0.528**
	-0.010		-0.025		0.046		0.037	
	(-0.15)	(1.42)	(-0.33)	(1.09)	(0.66)	(2.09)	(0.57)	(2.04)
$\text{RelFlow}_{i,t}$	-0.127**	-0.011*	-0.064	-0.021	-0.110*	-0.012**	-0.130**	-0.011*
	-0.030		-0.016		-0.026		-0.031	
	(-2.02)	(-1.9)	(-1.04)	(-0.89)	(-1.78)	(-2.08)	(-2.00)	(-1.91)
$\text{LnSize}_{i,t}$	0.000	-0.054	0.013	-0.056	0.014	-0.043	0.007	-0.051
	0.000		0.003		0.003		0.002	
	(0.00)	(-1.48)	(0.28)	(-1.51)	(0.34)	(-1.22)	(0.16)	(-1.41)
TurnRatio_t	0.001	0.000	0.002	0.000	0.001	0.000	0.001	-0.000
	0.000		0.000		0.000		0.000	
	(1.07)	(-0.33)	(1.34)	(0.46)	(1.42)	(0.13)	(0.94)	(-0.41)
$\text{LnAge}_{i,t}$	0.307**	0.128	0.346**	0.178*	-0.066	-0.122	0.262**	0.093
	0.074		0.084		-0.016		0.063	
	(2.95)	(1.35)	(2.84)	(1.89)	(-0.09)	(-0.21)	(2.43)	(0.97)
$\text{FundRisk}_{i,t}$	-6.948	-0.848	-6.772	-6.185	0.293***	0.120	-6.946	-0.990
	-1.668		-1.647		0.070		-1.667	
	(-1.07)	(-0.13)	(-0.92)	(-1.29)	(2.82)	(1.23)	(-1.07)	(-0.16)
YEAR FE	Y	Y	Y	Y	N	N	Y	Y
SEGMENT FE	Y	Y	Y	Y	Y	Y	Y	Y
N	5345	5345	3922	3922	5345	3922	5345	5345
Pseudo R-sq or R-sq	0.152	0.283	0.151	0.278	0.140	0.275	0.151	0.283

defined as follows:

$$\text{Stability}_{i,t} = -ln(1 + |\frac{\text{Dividend Ratio}_{i,t} - \text{Dividend Ratio}_{i,t-1}}{\text{Dividend Ratio}_{i,t-1}}|) \qquad (2.1)$$

Where Dividend Ratio is the sum of the dividend ratio, defined as distribution amount over reinvestment price, for each dividend distribution during the whole year. *Stability* is 0 when the dividend ratio in year t is the same as the dividend ratio in year t-1, and smaller when the difference between two variables, either positive or negative, is larger. Therefore, *Stability* is positively associated with dividend ratio stability. A mutual fund having more stable dividend ratios would get a higher value in *Stability*. The coefficients of *Stability* are positive and significant in columns (1) to (8). The results imply that dividend ratio stability is negatively associated with multi-dividend dummy and dividend frequency. It is consistent with the prediction of Hypothesis 2A.

In Hypothesis 2B, I predict that multi-dividend funds have lower expense ratio. Contrary to this prediction, the coefficients on *ExpRatio* are insignificant in columns (7) and (8), suggesting that expense ratio is not statistically different across dividend policy. Multi-dividend funds do not lower the expense ratio, a way that benefits investors, to collect sufficient dividend income.

We now turn our attention to another question: Do mutual funds trade off the interest between existing and new investors to attract new money? As denoted is section 3.2, in the segments with higher average dividend ratios, funds tend to pay dividends more frequently. Multivariate analysis also provides strong evidence at the fund level: the coefficients on *LnDivratio* are positive and significant in all columns. This shows that multi-dividend dummy and dividend frequency are positively associated with dividend ratio after controlling for other fund attributes.

I also find limited evidence supporting Hypothesis 4. The coefficients of risk-free interest rate in columns (5) and (6) are negative, though not significant in column (6), given other fund characteristics. The results suggest that funds are more likely to increase the dividend frequency when the interest rates go lower. It is consistent with the competition within equity funds.

Finally, we examine the hypothesis that mutual funds trade off interests within the existing investors with different dividend preferences. I use 12B-1 fees plus one fourth of front loads, as proposed by Huang et al. (2007), to proxy for participation costs. The coefficients on *PartCost* are positive across the first four columns, suggesting that dividend frequency increases with participation costs, consistent with Hypothesis 5. The results are also consistent with our argument that mutual funds have to solve the conflict of interest within existing investors.

2.5.3 Reaction of Investors

In this section, I answer the last question whether and how investors respond to multi-dividend policy. The ultimate objective for mutual funds is to increase the assets under management. Mutual funds choose the dividend policy that could maximize their assets. Therefore, I expect that multi-dividend policy increases the net inflows controlling for other fund characteristics. To test this hypothesis, I regress yearly mutual fund net inflows on multi-dividend dummy, fund characteristics including yearly dividend ratio, and their interaction terms following previous literature (e.g., (Barber et al., 2005; Kumar et al., 2012))[16]. Table 2.6 reports the results. The first column (left) in each specification reports the main effects of fund attributes on mutual fund relative flows. The right column in each regression reports coefficients and t-statistics on interaction terms between *MultiDiv* and fund attributes. It describes how multi-dividend policy alters the main effect coefficients in the left column.

We first investigate the main effects of *LnDivRatio* on the net mutual fund flows. The coefficients on *LnDivRatio* are negative, suggesting that mutual funds with high dividend ratios have smaller net flows than those with low dividend ratios. It shows that investors on average avoid investing in the funds with high dividend overhang. This result is consistent with Graham and Kumar (2006). They find that retail investors, as a group, prefer non-dividend-paying stocks over dividend-paying stocks.

Our primary focus is the relationship between dividend policy and fund flow. We would concentrate on the interaction term between fund characteristics one period lag behind and *MultiDiv*. The coefficients on interaction term *MultiDiv-LnDivRatio* are significant in the columns (1) and (2). It shows that multi-dividend policy positively alters the sensitivity of flows to dividend ratio. The absolute value of the coefficients on the interaction term is much larger than that of the coefficients on *LnDivRatio*. It suggests that dividend ratio is negatively associated with net fund flows in single-dividend funds whereas positively in multi-dividend funds. It is consistent with our hypothesis that multi-dividend policy reduces the dividend overhang problem and attracts more new investors. However, the sensitivity of new flow to risk-free interest rate is not significant in column (3), which contradicts our prediction.

Then we examine how investors react to the trade-off within investors with different dividend preferences. The regression reports that the coefficients on

[16] I exclude the variable *Stability* in this regression since it greatly decreases the sample size. Yet, even if I include that variable, the main results remain the same and the coefficients on that variable are not significant.

Table 2.6: Multi-Dividend Policy and Mutual Fund Flows

This table reports coefficients and associated t-statistics from OLS regressions of yearly relative flows on multi-dividend dummy, other fund characteristics, and their interaction terms. The dependent variables are yearly relative flows. The first column in each regression reports the main effects of fund attributes and multi-dividend dummy. The second column reports the interaction effects. All variables are defined in Table 2.2. Pquintile1 is the lowest performance quintile, defined as min(Prank, 0.2), where Prank is a fund's percentile performance relative to other funds in the same segment. It ranges from 0 (the worst funds) to 1 (the best funds). Pquintile2_4 is the second to fourth performance quintile, defined as min (Prank-Pquintile1, 0.6). Pquintile5 is the highest performance quintile, estimated as Prank-Pquintile1-Pquintile2_4. RelFlow is the yearly relative flow after winsorizing at 2% level. FundRisk is the standard deviation of raw returns during one year. ManFlow is the yearly relative net flows to fund *i*'s family. SegFlow is the yearly relative net flows to fund *i*'s segment. All other variables are defined in Table 2.2. Year FE, and Segment FE show whether we include year, and segment fixed effects. Standard errors are clustered by fund and year. *** indicates significance at 1% level, ** indicates significance at 5% level, and * indicates significance at 10% level.

	Yearly Relative Flow$_t$					
	(1)		(2)		(3)	
	Main Effect	Interaction	Main Effect	Interaction	Main Effect	Interaction
MultiDiv$_{i,t}$	-0.063		-0.087		-0.084	
	(-1.05)		(-1.35)		(-1.13)	
LnDivRatio$_{i,t-1}$	-0.660	1.797***	-0.912	1.948***	-1.186	2.047***
	(-1.17)	(2.96)	(-1.52)	(3.27)	(-1.95)	(3.36)
Fundret$_{i,t}$	0.007***	0.001***				
	(5.59)	(3.31)				
Pquintile1$_{i,t}$			0.053	0.005	0.066	-0.014
			(0.57)	(0.03)	(0.66)	(-0.07)
Pquintile2_4$_{i,t}$			0.094***	0.041	0.096***	0.040
			(4.38)	(1.41)	(4.68)	(1.2)
Pquintile5$_{i,t}$			0.540***	-0.362***	0.542***	-0.364***
			(5.79)	(-3.03)	(5.61)	(-2.9)
PartCost$_{i,t-1}$	-0.020***	0.023**	-0.019***	0.021*	-0.022***	0.020*
	(-2.63)	(1.96)	(-2.6)	(1.78)	(-2.91)	(1.66)
BackLoad$_{i,t-1}$	-0.05136	0.03655	-0.05535	0.048018	-0.04431	0.055293
	(-4.04)	(0.95)	(-4.42)	(1.17)	(-4.06)	(1.28)
No12B$_{i,t-1}$	-0.050 ***	0.000	-0.053***	0.004	-0.053***	0.008
	(-3.41)	(0.01)	(-3.7)	(0.22)	(-3.63)	(0.23)
LnSize$_{i,t-1}$	-0.015***	0.010***	-0.016***	0.010***	-0.016***	0.011***
	(-6.58)	(4.52)	(-6.23)	(4.53)	(-6.71)	(4.64)
TurnRatio$_{i,t-1}$	-0.000**	0.000	-0.000**	0.000	-0.000**	0.000
	(-2.26)	(0.96)	(-2.17)	(1.26)	(-2.32)	(1.33)
LnAge$_{i,t-1}$	-0.015***	-0.007	-0.017***	-0.005	-0.016***	-0.005
	(-2.83)	(-0.7)	(-2.75)	(-0.51)	(-2.65)	(-0.51)
FundRisk$_{i,t-1}$	-1.091*	0.070	-1.708***	0.138	-0.462	0.199
	(-1.91)	(0.18)	(-3.04)	(0.42)	(-1.36)	(0.31)
RF$_{i,t}$					-0.046	-0.035
					(-0.88)	(-0.44)
RelFlow$_{i,t-1}$	0.318***	0.028	0.315***	0.029	0.318***	0.027006
	(8.82)	(0.77)	(8.54)	(0.81)	(8.51)	(0.8)
ManFlow$_{i,t}$	0.002***		0.002***		0.002***	
	(2.6)		(2.39)		(2.28)	
SegFlow$_{i,t}$	0.126*		0.261***		0.398***	
	(1.79)		(2.78)		(7.52)	
Year FE	Yes		Yes		No	
Segment FE	Yes		Yes		Yes	
N	9362		9362		9362	
R^2	0.245		0.243		0.216	

MultiDiv-PartCost is positive and significant. It shows that a multi-dividend fund has a smaller outflow caused by the participation costs than a single-dividend fund, which supports our hypothesis 5. The results is also consistent with the hypothesis that dividend clientele prefer dividends in the purpose of obtaining incomes for consumption.

2.6 Conclusions

Corporate dividend policy has been extensively investigated in the literature. However, mutual fund dividend policy, which is quite different from corporate dividend policy, draws little attention in the literature. This chapter asks the question how and why mutual funds determine their dividend policy.

This chapter presents evidence that multi-dividend policy hurts investors and proposes that multi-dividend policy is the outcome of mutual funds optimally balancing two different conflicts of interest. The first conflict stems from the asymmetry of tax burdens between existing and new investors. New investors have preference to buy funds with low dividend overhang. To attract new investors, mutual funds have incentives to use multi-dividend policy in order to decrease dividend overhang. The second conflict of interest stems from the different dividend preferences of existing investors. Some investors might prefer high frequent distributions whereas some do not. Mutual funds have to trade off their interest when choosing dividend policy. To examine the hypotheses, I first relate the dummy for multi-dividend policy and dividend frequency to fund characteristics. I find dividend policy is associated with fund characteristics that affect both conflicts of interest. Then we regress fund net flows on fund characteristics, multi-dividend dummy, and their interaction terms. I find empirical evidence that multi-dividend policy attracts net flows related to dividend ratio, a variable associated to the conflict of interest between new and existing investors. The results show that multi-dividend policy is positively associated with net flows related to participation cost, a variable that measures the severity of the conflict of interest within investors. This result implies that some investors purchase mutual funds for the purpose of generating constant dividend income.

This chapter starts a new dimension in mutual fund analysis. We relate an important variation among mutual funds, dividend policy, to investor purchase decision. The two trade-offs documented in this chapter broaden two different areas. The first trade-off is associated with the agency problem that some mutual funds might not favor existing investors but new investors in order to maximize their benefits. The second trade-off, on the opposite, represents a delegation of investment decisions. Both trade-offs stem from the managers' desire to maximize the asset size under their management. This study helps us to understand managers' behaviors on dividend policy and redesign a better mutual fund fee structure.

This chapter also contributes to a large body of literature in corporate dividend policy. First, we provide a new and different type of dividend policy. Mutual fund dividend policy is highly regulated and different from the one of other corporates. Therefore, it is possible to test the hypotheses in payout policy literature while isolating some other factors. For example, DeAngelo and DeAngelo (2006) argue that Miller and Modigliani (1961) imposes an assumption to mandate 100% free cash flow payout in every period. It is difficult to find such corporates that pay out all free cash flows in empirical tests. However, mutual funds naturally fulfill this requirement. Second, we first provide evidence on investment purposes of dividend clientele. Previous literature suspects dividend clientele demands high dividends for financing their consumption or they have lower tax rates (Becker et al., 2011). Yet the real reason is difficult to test since some group of investors (e.g., older and low-income investors) have both characteristics (Becker et al., 2011; Graham and Kumar, 2006). We present evidence on that some investors chase dividends for the propose of consumption by showing that some investors react to dividend policy, which tax rationale cannot explain.

3

Who Buys the Worst Mutual Funds? Fund Performance and Investor Characteristics

Contents

3.1 Introduction

The vast majority of assets under management are invested in actively managed mutual funds despite the abundant empirical evidence that actively managed equity mutual funds, on average, underperform passively managed mutual funds after fees and transaction costs (Gruber, 1996). Moreover, mutual funds that outperform cannot be identified *ex-ante* based on past performance (Carhart, 1997) and true skill is very scarce (Fama and French, 2010). Funds with negative alpha, however, are not rare and can be identified both from past performance and fund attributes, such as high fees (Carhart, 1997). Some authors argue that this puzzling evidence can be explained by the presence of financially unsophisticated investors in the market for mutual funds. Gruber (1996) was the first to talk about a "disadvantaged" mutual fund clientele. Despite the appeal of the argument, identifying the disadvantaged clientele is a difficult task because of lack of comprehensive data on investors' mutual fund investment decisions. The available data sets, discussed below, are typically obtained from micro data from Nordic countries, surveys, and individual account data provided by a broker. However, it is unclear whether the conclusions obtained from those datasets can be extrapolated to the whole US market, with distinct institutional and cultural characteristics and a different competitive structure, and a more heterogeneous investor population than the subset of investors that operate through one broker. The purpose of this chapter is to identify the sociodemographic characteristics of investors investing in the worst mutual funds using a novel data set of US Internet visitors to websites of mutual fund management companies.

From the point of view of policy design, it is useful to know which groups of investors, as defined by their sociodemographic characteristics, are more likely to be financially unsophisticated. This information can allow regulators to target policies aimed at enhancing investor protection to the most vulnerable groups. It is also interesting to know whether and how poor financial decision making contributes to observed wealth inequalities. This question is particularly relevant if we consider the important role of mutual funds as an investment vehicle and the potential long-term wealth loss arising from poor investment decisions.[1] Finally, exploring how heterogeneity across investors is related to differences in financial decision making can shed light on the reasons why retail investors often fail to make optimal financial decisions.

[1] 44.1% households in the US hold mutual fund shares according to the Investment Company Institute (http://www.ici.org). Moreover, mutual fund investment accounts for 54% and 47% of all investment in Defined Contribution Plans and IRAs, respectively.

Although the existence of a sufficiently large disadvantaged clientele may make it possible for underperforming funds to survive, the survival of underperforming funds by itself does not imply that there is heterogeneity in mutual fund investor sophistication. For instance, all investors could be equally likely to purchase underperforming funds if there were costs to acquiring and processing the relevant information and those costs were similar across investors. To investigate whether some investors are more likely than others to choose underperforming funds, the literature has followed two different paths. The first one looks for indirect evidence that the market for mutual funds is segmented. Building on Gruber's (1996) insight, Gil-Bazo and Ruiz-Verdú (2008) show theoretically that a negative cross-sectional relation between fees and performance could emerge in a market with asymmetric information about managerial skill if a fraction of investors have a less than perfectly elastic demand with respect to fund net performance. These authors show that in equilibrium poor-quality funds cater to unsophisticated investors exclusively and charge higher fees. Consistently with this prediction, Gil-Bazo and Ruiz-Verdú (2009) report evidence of a negative relation between fees and before-fee performance in the cross-section of US equity funds. The idea that heterogeneity in financial sophistication may lead to segmentation of the mutual fund market is also supported by the empirical evidence that funds distributed through the broker channel underperform their peers in the direct channel even before sales charges (Bergstresser et al., 2009; Christoffersen et al., 2013; Del Guercio and Reuter, 2011).

Another strand of the literature has used surveys, experiments and micro data to study directly the optimality of investors' decisions. Alexander et al. (1998) document that only 16% of survey respondents holding mutual fund shares understand that higher expenses might result in lower average returns. Choi et al. (2010) report that a majority of subjects participating in an experiment fail to minimize fees when choosing among otherwise identical funds. A number of studies have also explored the relation between investors' mutual fund choices and their degree of financial literacy, proneness to behavioral biases, and intelligence. Hastings and Tejeda-Ashton (2008) report that individuals that experience difficulties in providing correct answers to basic financial questions are less sensitive to mutual fund fees. Using data on individual investors' trades and end-of-month portfolio holdings from a large discount brokerage house, Bailey et al. (2011) report that investors that are more prone to behavioral biases, as inferred from their stock trades, are also more likely to make bad mutual fund choices resulting in poorer performance. Finally, Grinblatt et al. (2011) link a comprehensive data set on Finnish males' IQ scores to fund holdings and find that IQ, which the authors view as a proxy for the cognitive costs of search, is negatively associated with fund fees, controlling for many other fund and investor characteristics.

As mentioned above, regulators may find it useful to know which groups of investors are more likely to be financially unsophisticated. While financial literacy, behavioral biases, or IQ, are not observable to regulators, sociodemographic characteristics are. A number of studies, including those mentioned above, investigate the association between variables such as age, gender, marital status, income or education, and mutual fund investment decisions. Engström (2007) documents male investors chase funds with better past returns while male, high educated and more experienced investors choose high-fee funds. Niebling et al. (2009) report that older, more experienced and wealthier investors are more likely to follow past performance. Tang et al. (2010) show that younger and higher-income participants make better investment decisions. Bailey et al. (2011) find that older and more experienced investors are less likely to invest in high-fee funds. They also show that male and young investors are more likely to chase past returns. Grinblatt et al. (2011) report that having a university or business degree, working in the finance profession, and wealth, are associated with choosing low-fee funds. Choi et al. (2010) do not find any evidence that high SAT scores are related to fees. Finally, exploiting international heterogeneity, Ferreira et al. (2012) document that the flow-performance relationship is less convex in the countries with higher education levels.

An important challenge in this literature is the limited availability of data on mutual fund investor characteristics, which is in contrast with the large amount of information on mutual funds (Campbell, 2006; Thaler, 1999). Table 3.1 describes the three main sources of data employed in the literature. The first source of data are micro data from Nordic countries. Engström (2007) obtains investor characteristics from *Statistics Sweden*. His database covers 147,000 individuals, representing 3.5% of the Sweden population. It provides various sociodemographic characteristics, such as age, gender, education, income and marital status. Investor characteristics in Grinblatt et al. (2011) come from two sources. IQ scores are provided by *Finnish Armed Forces*. The other characteristics, such as education, occupation, are from *Statistics Finland*. Their sample includes about 7,500 male subjects per year who hold funds. The second source of data are surveys. Bauer and Smeets (2010) gather investor characteristic data from the clientele of two socially responsible banks in the Netherlands. Their sample includes 3,187 individual investors and 22 mutual funds. Alexander et al. (1998) conduct a survey covering 2,000 distinct respondents from 6 different distribution channels in the US. Their investor characteristics include gender, age, education and income. The third source of data are actual transactions carried out through some broker. The investor characteristics in Bailey et al. (2011) are from a US discount broker. Their final sample consists of 1,492 different equity mutual funds and 29,381 investors. Niebling et al. (2009) use a dataset containing

46,000 distinct individual mutual fund investors provided by a German discount brokerage house. The investor characteristics in both papers incorporate gender, age, income/wealth and marital status. Tang et al. (2010) collect individual attributes of nearly one million investors from Vanguard, a leading 401(k) administrator and mutual fund investment manager. Their investor characteristics include age, sex, plan tenure, non-retirement financial wealth, household income, home ownership status, and whether the participant had web access.

Table 3.1: Data Sources for Investor Characteristics Employed in the Literature

Study	Data Source	Source Type
Grinblatt et al. (2011)	Finnish Armed Forces IQ test Statistics Finland	Micro data
Engström (2007)	Statistics Sweden	Micro data
Bauer and Smeets (2010)	Survey in the Netherlands	Survey
Alexander et al. (1998)	Survey of US fund investors	Survey
Bailey et al. (2011)	Transactions from a US broker from 1991 to 1996	Sample of Investor Transactions
Tang et al. (2010)	401(K) pension plans in Vanguard	Sample of Investor Transactions
Niebling et al. (2009)	Transactions from a German broker	Sample of Investor Transactions

We contribute to the literature by investigating the relation between mutual fund investor characteristics and investment decisions using a novel data set of the sociodemographic characteristics of US Internet visitors to mutual fund websites. We obtain the data set, which is described in detail in Section 3.2, from Quantcast, an Internet audience measurer, which provides information on the sociodemographic composition of visitors to Internet websites. Our data set has a number of appealing features. First, it captures the characteristics of nearly all investors who search information through the Internet. We argue below that there is a substantial overlap between mutual fund investors and Internet visitors to mutual fund websites, which makes our sample representative of the investor population. Second, the data correspond to the world's largest and most-studied market for mutual funds, the US market, which accounts for approximately half of all worldwide mutual fund assets. At the end of 2011, assets under management in the Finnish and Sweedish mutual fund markets were 0.53% and 1.54% of assets under management in the US market, respectively.[2] Even abstracting from size, it is an open empirical question whether results found for those countries can be extrapolated to other countries due to international differences in household characteristics, culture, institutional frameworks, and competitive structures (Bover, 2010; Bover et al., 2013; Cremers et al., 2011; Ferreira et al., 2012). Engström and Westerberg (2004) and Engström (2007) document that investors in Sweden behave differently from investors in the US. Third, our data set contains some sociodemographic char-

[2]Data from the Investment Company Institute.

acteristics which have not been studied before in the literature on mutual fund choices, such as ethnicity. Fourth, the data are up-to-date and freely available. In contrast, the data set recently used by Bailey et al. (2011) is from the 1991-1996 period. It is unclear whether investor behavior has experienced any substantial changes since then. Finally, as opposed to data provided by one broker, our database covers investors in both the direct and indirect channels. Del Guercio et al. (2010) argue that funds in the direct and indirect channels target different types of investors. Arguably, the decision to operate through a specific broker depends on investor characteristics, which makes a broker-provided sample not representative of the whole population of mutual fund investors.

Of course, the data set we employ in this chapter is not caveat-free. First, investor characteristics must be necessarily aggregated at the management company level since websites are maintained for fund management companies, not individual mutual funds. Second, investor sociodemographic characteristics are approximated from those of Internet visitors, which implicitly assumes the representativeness of the Internet visitors' sample. Finally, our data set is a single cross section. Despite these drawbacks, given the scarcity of data on investor characteristics we believe that our data set is an interesting alternative to study the determinants of retail investor decisions.

In our analysis, we start by defining the worst funds as those with poor forecasted alphas. To estimate forecasted alpha, we follow the literature and regress measures of risk-adjusted performance on fund characteristics that have been documented to predict performance. Then, we use a logit model to examine the association between investor sociodemographic characteristics and the probability that the fund's forecasted alpha belongs to the bottom of the distribution.

Our main results can be summarized as follows. First, we find that mutual funds with a higher fraction of female, older, and low-income investors have a higher probability of belonging to the group of worst funds. We also document that funds with a higher fraction of African American investors are associated with better predicted performance. Investor education, however, does not appear to be associated with the optimality of investment decisions. Second, we document that the inferior decisions of some groups of investors result mainly from a failure to react to poor past performance and fund fees. Third, we investigate whether the decision to invest in underperforming funds by some groups of investors is induced by a higher sensitivity to fund marketing, i.e., brokers' advice and advertising. We do not find evidence supporting that hypothesis with a single exception: the oldest investors.

The rest of the chapter is organized as follows. Section 3.2 provides the background of our data, the database description and summary statistics. Section 3.3 describes the empirical strategy and presents the chapter's results. Section 3.4 concludes.

3.2 Data

3.2.1 Investor Characteristics

Quantcast estimates the sociodemographic characteristics of visitors to Internet websites by combining two different sources of data: Directly Measured Data (DMD) and Multiple Reference Data (MRD). Quantcast collects comprehensive census records of media consumption from 80,000 partner publishers covering 10 million website destinations each month, including websites, blogs, videos, widgets, and advertising campaigns. DMD represent the media consumption activity of over 200 million people in the US. Quantcast also collects Multiple Reference Data (MRD), such as click-stream and non-personally identifiable information user data, from market research companies, Internet service providers (ISPs) and toolbar vendors covering over 1.5 million users in the US. The methodology employed by Quantcast extrapolates information on investor characteristics obtained from MRD to all the covered websites by using a vast amount of data on visitors' activity contained in DMD.[3]

Internet visitor characteristics are a good proxy for investor characteristics only to the extent that sociodemographic characteristics are distributed similarly in both groups. There are reasons to believe that the sets of fund investors and mutual fund website visitors largely overlap. The Internet has become one of the most important tools for investors, especially for retail investors, to obtain information on mutual funds. Da et al. (2011) provide evidence that Internet search activity correlates with retail investor attention. In a 2010 survey of mutual fund investors conducted by the Investment Company Institute (ICI), almost all respondents report that they have Internet access at home and 82% of investors use the Internet for financial purposes. Among households with mutual funds, 79% access their financial accounts through Internet; 58% use Internet to get investing information; 21% of mutual fund households buy mutual funds directly through Internet (Bogdan et al., 2010). Other sources of data are indicative that visitors to mutual fund family websites invest with

[3] See *Quantcast Methodology Overview*, available at http://www.quantcast.com/white-papers/quantcast-methodology.pdf, for further details.

those mutual fund families. For example, according to the Internet traffic company, Compete, during January 2011, *www.fidelity.com*, the official website of Fidelity mutual funds management company, was visited by 5,647,728 unique visitors, among which, 4,656,329, reached *login.fidelity.com.* That is, at least 82% of Fidelity web visitors hold investment accounts with Fidelity.[4] Below, we document that the distribution of sociodemographic characteristics is similar in both our data of mutual fund website visitors and mutual fund investors, while the characteristics of mutual fund website visitors are different from those of Internet users.

Admittedly, our database suffers from self-selection, as the decision to operate through the Internet is likely to be endogenous to investor characteristics. However, self-selection does not necessarily invalidate inference about differences in investor characteristics between unsophisticated investors (those who purchase bad funds) and sophisticated investors based on the Internet sample. It could be argued that usage of Internet depends on investor sophistication but not on other investor characteristics. In that case, the distribution of investor sociodemographic characteristics among both the unsophisticated and the sophisticated groups remains unaltered in the Internet sample relative to the general population. If the probability that an investor operates through the Internet depends on the investor's sociodemographic characteristics but not on sophistication, certain categories of investors will be overrepresented in the Internet sample relative to the entire population, which will bias the distribution of sociodemographic characteristics among both unsophisticated and sophisticated investors. Under the null hypothesis that sociodemographic characteristics have the same distribution within both types of investors, the bias in the distribution is identical for both, so the null holds true for the Internet sample, too. The only cause for concern is that the probability of operating through the Internet depends on both sophistication and investor characteristics. In this paper, we implicitly assume that Internet usage depends on either investor sociodemographic characteristics or sophistication but not on both.

We keep the largest 250 management companies in terms of assets under management in June 2010. We drop the smallest mutual fund management companies because information on their websites is typically unavailable or noisy due to the small number of visitors. We employ information from CRSP and search engines (Google or Bing) to obtain web addresses of mutual fund management companies. If there are several websites or addresses pointing to the same management company, we choose the one with largest number of clicks. For example, there are two domains pointing to Columbia management company: *www.columbiafunds.com* and *www.excelsiorfunds.com*. Yet, we use the former since it has more clicks than the latter: According to Quantcast,

[4]The data is downloadable from *www.compete.com*

www.columbiafunds.com has 12,800 visitors per month while so few clicks are recorded on *www.excelsiorfunds.com* that Quantcast cannot give any information.

We input those websites into Quantcast to obtain investor characteristics. Quantcast offers data on the composition of website visitors in terms of various sociodemographic characteristics in six categories: gender, age, ethnicity, parenthood, income and education. Table 3.2 shows definitions of the variables for investor characteristics. We are able to obtain statistics for 207 management companies as of June 2010. We have verified that sociodemographic characteristics of Internet visitors exhibit almost no variation through time. Therefore, we perform our analysis on a single cross-section.

Table 3.2: Definition of Variables for Investor Sociodemographic Characteristics

Category	Variable	Definition
Gender	Male	Percentage of male visitors
	Female	Percentage of female visitors
Age	3-12	Percentage of visitors from 3 to 12 years old
	13-17	Percentage of visitors from 13 to 17 years old
	18-34	Percentage of visitors from 18 to 34 years old
	35-49	Percentage of visitors from 35 to 49 years old
	50+	Percentage of visitors above 50 years old
Ethnicity	Cauc	Percentage of Caucasian visitors
	AfrAm	Percentage of African American visitors
	Asian	Percentage of Asian visitors
	Hisp	Percentage of Hispanic visitors
	Other	Percentage of other visitors
Children	Has Kids0-17	Percentage of visitors with 0-17 year-old children
	No Kids0-17	Percentage of visitors without 0-17 year-old children
Income	0K-30K	Percentage of visitors with income of 0-30 thousand dollars per year
	30K-60K	Percentage of visitors with income of 30-60 thousand dollars per year
	60K-100K	Percentage of visitors with income of 60-100 thousand dollars per year
	100K+	Percentage of visitors with income of more than 100 thousand dollars per year
Education	No College	Percentage of visitors without entering college
	College	Percentage of visitors with college degree
	Postgrad	Percentage of visitors with postgraduate degree

Table 3.3 compares the characteristics of mutual fund website visitors in our database with those of all Internet users and actual mutual fund investors. Columns (1), (2) and (3) report the mean, median and asset-weighted mean of website visitors, respectively. Column (4) shows the characteristics of all Internet users, as estimated by Quantcast. Columns (5) and (6) report the composition of Internet users older than 3 years and 16 years, respectively, as reported by the National Telecommunications & Information Administration (NTIA). Column (7) presents the compositions of investors as reported by the Investment Company Institute.

Considering the differences in categorization across databases, we find that the distribution of individual characteristics in our sample is very similar to that of

Table 3.3: Characteristics of Mutual Fund Website Visitors, Internet Users, and Mutual Fund Investors
The table shows the summary statistics of the percentages of investors belonging to different sociodemographic groups in the sample of mutual fund website visitors, the population of Internet users and the population of mutual fund investors. Variables are defined in Table 2. The database of mutual fund website visitors, columns (1), (2) and (3), is obtained for June, 2010. Internet population characteristics in column (4) are obtained for June, 2012. The Internet population characteristics, columns (5) and (6), are provided by the National Telecommunications & Information Administration (NTIA) for 2010. The two columns (5) and (6) correspond to the population older than 3 years and 16 years, respectively. Mutual fund investor characteristics, column (7), are reported by the Investment Company Institute (ICI) for 2010. Figures have been adjusted in order to make the categories homogeneous across columns. In column (1), standard deviations are given in parentheses.

	Mutual Fund Website Visitors			Internet Users			Mutual Fund Investors
	Mean (1)	Median (2)	Asset-Weighted (3)	Quantcast (4)	NTIA 2010(3+) (5)	NTIA 2010(16+) (6)	ICI (7)
Male	57.80 (9.36)	59	57.79	49	48.93	48.44	18
Female	42.20 (9.36)	41	42.2	51	51.07	51.56	20
Co-decided							62
3-12	2.84 (2.22)	3	1.32				
13-17	6.30 (4.73)	5	3.08				
0-18				16	20.97		
18-34	20.68 (8.53)	19.5	19.06	29	24.45	33.14*	15
35-49	29.84 (8.69)	30	30.92	27.5	21.27	26.06	33.5
50+	40.41 (13.59)	40	45.69	24.5	33.3	40.8	51.5
Cauc	80.20 (9.51)	81	83.17	76	65.14	67.49	90
AfrAm	6.99 (7.25)	5	5.45	9	12.01	11.59	5
Asian	6.62 (6.30)	4	5.68	4	4.59	4.65	2
Hisp.	5.04 (3.45)	5	4.73	10	15.87	14.26	5
Other	1.11 (0.65)	1	0.83	1	0.5	0.59	4
No Kids0-17	76.89 (8.73)	78	80.71	59.36			
Has Kids0-17	23.11 (8.73)	22	19.29	41.47			
0K-30K	16.58 (7.54)	15	13.64	18	30.27	29.96	11
30K-60K	27.48 (7.93)	26	25.66	26	27.46	27.88	24
60K-100K	28.79 (7.54)	29	30.16	28	23.1	23.26	31
100K+	27.11 (8.69)	26.5	30.67	28	19.17	18.9	36
No College	33.65 (8.81)	33	35.06	45			40
College	43.80 (7.47)	43	42.98	41			34
Postgrad	22.59 (8.25)	22	22.03	14			26

*From 16 to 34 years old.

all mutual fund investors.[5] Also, individuals in our database, just like mutual fund investors, do not appear to be a representative sample of all Internet users. For example, our data shows that 45.69% of visitors to mutual fund websites (asset-weighted) are older than 50 years. This figure is close to the fraction of all mutual fund investors that belong to the same age group, 51.5%, but higher than the fraction of Internet users in the same category, between one quarter and one third. Similarly, mutual fund website visitors in the lowest-income group are only 13.64%, which is consistent with 11% of all mutual fund investors reported to belong to that group, and much lower than 30%, which is the fraction of all low-income Internet users, as reported by NTIA. The data are in line with those of Bogdan et al. (2010), who document that older, poorer and lower educated investors are associated with a lower level of Internet usage. The results of Table 3.3 provide support to the idea that the sample of mutual fund website visitors is representative of the population of mutual fund investors.

3.2.2 Mutual Fund Performance

We obtain mutual fund data from the CRSP Survivor-Bias Free Mutual Fund Database spanning from January 1995 to June 2010. The dataset includes all active mutual funds during that period. Following the previous literature (Carhart, 1997; Chen et al., 2004; Gil-Bazo and Ruiz-Verdú, 2009), we keep domestic equity mutual funds as defined by their investment objectives codes. We also exclude index, and institutional funds identified by both fund names and CRSP identifiers. We use fund names to identify funds with multiple share classes and compute for each fund the asset-weighted mean of each of the following fund characteristics: return, expense ratio, management fee, 12b-1 fee, front and rear load. Fund age is calculated as the difference in months between the first offer date of the oldest class in the portfolio and the current date. We sum total net assets for all classes in the portfolio. Following Elton et al. (2011) and Evans (2010), we drop funds with total net assets below $15 million and age less than 36 months. Our final dataset consists of 284,371 fund-month observations corresponding to 3,754 distinct funds across 204 months. For each month, we have an average of 1,394 funds and 357 management companies.

[5] Categorizations of different investor characteristics are not exactly the same across databases. For example, Quantcast uses three categories for education: “No College”, “College” and “Postgraduate”, whereas ICI uses 4 categories: “High school graduate or less”, “Associate’s degree or some college”, “Completed college” and “Some graduate school or completed graduate school.” To homogenize categories, we assume that half of the ”Associate’s degree or some college” is categorized as ”College” in Quantcast data while the other half is corresponding to ”No college”.

We estimate risk-adjusted performance, alpha, using the four-factor model proposed by Carhart (1997):

$$r_{i,t} = \alpha_i + \beta_{rm,i} rm_t + \beta_{smb,i} smb_t + \beta_{hml,i} hml_t + \beta_{mom,i} mom_t + \varepsilon_{i,t} \quad (3.1)$$

Where $r_{i,t}$ is the net return in month t in excess of the one month T-bill return; α_i is the fund's alpha and captures the fund's risk-adjusted performance; rm_t is the market portfolio return in excess of the risk-free rate; smb_t is the return on a portfolio of small stocks minus large stock; hml_t is the return on a portfolio long high book-to-market stocks and short low book-to-market stocks; mom_t is the return difference between stocks with high and low past return in the past 12 months.[6]

We run (3.1) for each month t (from January, 1998 to June, 2008) using data from the prior 3 years (from $t-36$ to $t-1$) . We require at least 30 non-missing observations over the prior 36 months. We then define the realized alpha for fund i at period t, $\alpha_{i,t}$, as the fund's after-fee return minus the realized risk premium:

$$\alpha_{i,t} \equiv r_{i,t} - (\hat{\beta}_{rm,i} rm_t + \hat{\beta}_{smb,i} smb_t + \hat{\beta}_{hml,i} hml_t + \hat{\beta}_{mom,i} mom_t) \quad (3.2)$$

We are interested in identifying mutual funds that can be predicted to underperform. To obtain a proxy for future forecasted alpha, we first regress realized alpha on lagged fund characteristics that have been documented to predict future performance using monthly data from June 1998 to June 2008. In particular, we follow Chen et al. (2004) and regress realized alpha on fund size, family size, fund age, expense ratio, turnover ratio, and past returns in the previous 12 months:

$$\begin{aligned} \alpha_{i,t} = \mu_i &+ \gamma_1 LnTNA_{t-1} + \gamma_2 LogFsize_{t-1} \\ &+ \gamma_3 Turnover_{t-1} + \gamma_4 LnAge_{t-1} + \gamma_5 Expratio_{t-1} \\ &+ \gamma_6 Totload_{t-1} + \gamma_7 Flow_{t-1} + \gamma_8 FundRet_{t-1} + \varepsilon_{i,t}, \end{aligned} \quad (3.3)$$

where $LnTNA$, $LogFsize$, $Turnover$, $LnAge$, $Expratio$, $Totload$, $Flow$ and $FundRet$ are defined in Table 3.4.

We use both the Fama-MacBeth two-stage regressions and pooled OLS to estimate (3.3). Table 3.5 reports the results. Columns (1) and (2) show the coefficients and t-statistics (in parentheses) using pooled OLS and Fama-MacBeth methods, respectively. Results are similar for both methods. Because not all

[6] Data are obtained from Kenneth French's Website.

Table 3.4: Definition of Variables for Mutual Fund Characteristics

Variable	Definition	Units
TNA	Total net asset under management of the portfolio	million
LnTNA	Natural logarithm of TNA under management of the portfolio	million
LogFsize	Natural logarithm of total net asset under management of the mutual fund company	million
Turnover	Minimum of aggregate sales or purchases divided by the average TNA over the past 12 months	% per year
LnAge	Natural logarithm of the differences between the date and the first offered date	months
Expratio	Total fund's operating expenses divided by the year-end TNA	%
Totload	Max front-end load charged plus the rear-end load for holding 48 months as a proportion of investment	%
Flow	New fund flow in this month divided by the fund lagged TNA	%
FundRet	Cumulative fund return over the past one year	%

variables are significantly associated with future performance, we rerun the regression using significant explanatory variables only. The results of the new model are shown in columns (3) and (4), respectively.

We then calculate predicted alpha, which we denote by $\widehat{\alpha}_{i,t+1}$, as the product of the estimated coefficients in equation (3.3) and the fund characteristics at time t. We transform the fund characteristics from panel data to cross-sectional data by averaging variables from July 2008 to June 2010. We require at least 12 non-missing values to compute the mean. Table 3.6 reports summary statistics for our dataset.

Finally, we merge investor characteristics with fund data. Although our mutual fund data contains information on 2,446 mutual funds, we have investor characteristic data for 193 mutual fund management companies managing 1,858 funds, which represent more than 90% assets of the mutual fund sample.

3.3 Empirical Strategy and Results

3.3.1 Empirical Strategy

Our empirical strategy has two parts. First, we define the group of "the worst funds" as those whose predicted alpha, is in the 15% or 20% bottom

Table 3.5: Fund Performance and Lagged Fund Characteristics
The table displays the results from monthly regressions of fund performance on lagged fund characteristics for each month from January, 1998 to June, 2008. Fund performance is calculated as realized alpha, i.e., fund return minus the product of betas and factor realizations using Carhart's four-factor model. Variables are defined in Table 4. Regressions are estimated using the Fama-MacBeth approach in columns (1) and (3) and pooled OLS with month fixed effects in columns (2) and (4). The t-statistics are adjusted using Newey-West (1987) with lags of order three in Fama-MacBeth and clustered by fund and time in pooled OLS regressions. The t-statistics are shown in the parentheses. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.The coefficients reported are 100 times larger than the real value.

	Realized Alpha			
	Pooled OLS	Fama-MacBeth	Pooled OLS	Fama-MacBeth
	(1)	(2)	(3)	(4)
$LnTNA_{t-1}$	-2.547**	-1.900**	-4.162**	-4.232 ***
	(-2.383)	(-2.160)	(-2.282)	(-4.593)
$LogFsize_{t-1}$	0.834*	0.53	0.2946	
	(1.737)	(1.136)	(0.248)	
$Turnover_{t-1}$	-0.036	-0.010		
	(-0.879)	(-0.261)		
$LnAge_{t-1}$	3.307***	2.370***	8.122 ***	7.06 ***
	(2.970)	(2.638)	(3.776)	(4.929)
$Expratio_{t-1}$	-7.815*	-7.250 **	-12.009**	-8.704**
	(-1.902)	(-1.988)	(-2.154)	(-2.136)
$Totload_{t-1}$	-1.351	-1.180		
	(-0.699)	(-0.674)		
$Flow_{t-1}$	0.517	0.550		
	(0.764)	(1.289)		
$FundRet_{t-1}$	0.995**	0.770***	4.087 ***	3.987***
	(2.226)	(2.773)	(13.854)	(18.109)
Intercept	-186.951***	-91.309***	-411.679***	-445.097***
	(-3.442)	(-3.057)	(-13.941)	(-14.805)
Month Fixed Effects	YES	NO	YES	NO
N	102748	102748	115316	115389
Adjusted R^2	0.071	0.081	0.222	0.241

Table 3.6: Summary Statistics of Fund Characteristics
The table shows the summary statistics of variables for fund characteristics. Predicted alpha (Pooled OLS) and Predicted alpha (Fama-MacBeth) is calculated as the product of the vector of fund characteristics at time t and the vector of estimated coefficients from pooled OLS and Fama-MacBeth regressions, respectively. 12b-1fee is the percentage of the total assets attributed to marketing and distribution costs. All other variables are defined in Table 4. Variables are time-series means computed over the period from June, 2008 to June, 2010.

	Obs	Mean	Std. Dev.	25%	Median	75%
Predicted alpha(Pooled OLS)	1254	0.006	1.059	-0.286	-0.136	0.036
Predicted alpha(Fama-MacBeth)	1254	-0.474	1.035	-0.764	-0.618	-0.446
FundRet	1405	99.842	27.502	92.214	95.354	99.505
Total load	2446	0.271	0.428	0.000	0.000	0.500
12b-1fee	2446	0.139	18.764	0.000	0.026	0.250
Expratio	1863	1.165	0.340	0.963	1.162	1.390
LnTNA	2446	5.248	1.584	3.988	5.106	6.332
LnAge	2446	4.870	0.669	4.447	4.870	5.254

of the cross-sectional distribution of predicted alphas, with predicted alpha calculated as the mean of the time series of $\widehat{\alpha}_{i,t+1}$ from July 2008 to June 2010.

We then estimate a logit model to investigate how the sociodemographic characteristics of investors are associated with fund quality. More specifically, we model the probability of a fund belonging to the underperformers' group as a function of the characteristics of the investors in the fund's family:

$$P(Y_i = 1) = \frac{1}{1 + exp(-\alpha - \beta X_i - \varepsilon_i)}, \tag{3.4}$$

Where Y_i is a dummy variable that equals 1 if fund i belongs to the "worst funds" group and 0 otherwise, and X_i is a vector of explanatory variables containing *Female, 35-49, 50+, AfrAm, Asian, Hisp, 30K-60K, 60K-100K, 100K+, College, Postgrad* as defined in Table 3.2.[7]

[7]We assume that all mutual fund investors are adults (older than 18 years). Therefore, we treat positive fractions of Internet visitors in age groups *3-12* and *13-17* in the Internet database as errors. We eliminate them and re-scale other variables in age category so that their sum is one for each fund. We also drop variables in the *Kid* category from investor characteristics because they are highly correlated with *Age*. *Male, 18-34, Cauc, 0K-30K, No College* serve as the omitted categories for gender, age, ethnicity, income and education, respectively. The correlation table is available from the authors upon request.

3.3.2 Who Buys the Worst Mutual Funds?

In Table 3.7, we report the chapter's main results. The table contains coefficients and marginal effects as well as test statistics (z-values), of the logit regression (3.4) with standard errors clustered at the management company level. The marginal effects and reference probabilities are calculated at the average values of regressors. For the sake of providing a more complete picture of the relation between investor characteristics and fund choices, we also estimate a logit model for the probability of the fund being in the group of funds with highest predicted alpha.

We start by analyzing how gender relates to fund choices. The coefficients and marginal effects for *Female* are positive and statistically significant across all four proxies of predicted alpha in columns (1) to (4). The marginal effects reveal that a one percent increase in the proportion of female investors increases the probability that a fund is in the worst performance group by 0.45%. This result suggests that the fraction of female investors is higher in funds expected to underperform. A possible explanation for this association is that female investors are, on average, less financially sophisticated than male investors. This conclusion is consistent with Lusardi and Mitchell (2011), who report that women are less financially literate than men. Also, Engström (2007) and Bailey et al. (2011) document that male investors are more likely to follow past mutual fund returns. If less return-chasing implies that females are less sensitive to performance, then female investors would face a higher risk of staying in funds that have underperformed in the past and are more likely to underperform in the future. The coefficients on *Female* reported in columns (5) to (8), corresponding to the probability of a fund belonging to the top-performing group, are negative although the coefficients in specifications (5) and (7) are not significant.

The coefficients and marginal effects for age group variables *35-49* and *50+* are positive and significant in columns (1) to (4). These results indicate that the fraction of middle-aged and older investors is higher in funds with lowest predicted alphas. A one percent increase in the proportion of investors in the *35-49* and *50+* age groups increases the probabilities of a fund being bad by 0.8% and 0.4%, respectively. Therefore, young investors (omitted category) appear to exhibit more ability to stay away from future underperformers. This finding is consistent with Bailey et al. (2011), who show that age is negatively related to return-chasing behavior and with Tang et al. (2010), who find that young investors experience smaller losses from inefficient investment choices.

The coefficient for *35-49* is positive but larger than that of *50+* in columns (1)

Table 3.7: Investor Characteristics and Predicted Performance
The table reports coefficients and marginal effects and associated z-values from a logit regression for the probability that the fund belongs to the group of funds with bottom/top predicted alpha. The first, second and third row of each variable reports coefficients, marginal effects and z-statistics, respectively. The dependent variable equals 1 if a fund's predicted alpha is at the bottom/top 15%/20%. Predicted alpha is calculated as the time-series mean of predicted alphas calculated by Fama-MacBeth (F-M) or pooled OLS (OLS) regressions from June, 2008 to June, 2010. All other variables are defined in Table 2. Standard errors are clustered by management company. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.

	Predicted Alpha							
	15% bottom (F-M) (1)	20% bottom (F-M) (2)	15% bottom (OLS) (3)	20% bottom (OLS) (4)	15% top (F-M) (5)	20% top (F-M) (6)	15% top (OLS) (7)	20% top (OLS) (8)
Female	3.728**	2.497*	3.524**	3.146**	-1.744	-3.699**	-1.895	-3.857***
	0.457**	0.394*	0.430**	0.494**	-0.215	-0.579**	-0.234	-0.603***
	(2.362)	(1.736)	(2.229)	(2.083)	(-1.219)	(-2.505)	(-1.291)	(-2.649)
35-49	6.428***	5.666***	6.452***	5.862***	-1.548	-1.231	-1.490	-1.596
	0.788***	0.893***	0.787***	0.921***	-0.191	-0.193	-0.184	-0.25
	(2.826)	(2.905)	(2.826)	(2.938)	(-0.940)	(-0.713)	(-0.883)	(-0.938)
50+	3.256**	2.496*	3.216**	2.676**	-1.615	-1.396	-1.445	-1.643
	0.399**	0.393*	0.392**	0.420**	-0.200	-0.218	-0.178	-0.257
	(2.00)	(1.877)	(1.964)	(2.019)	(-1.204)	(-1.080)	(-1.042)	(-1.314)
Afram	-3.745	-5.134*	-3.863	-5.788**	5.817***	4.412**	5.750***	4.596**
	-0.459	-0.809*	-0.471	-0.909**	0.718***	0.690**	0.709***	0.719**
	(-1.358)	(-1.952)	(-1.390)	(-2.039)	(3.119)	(2.197)	(3.034)	(2.325)
Asian	-2.869	-1.032	-2.807	-1.82	-1.977	-0.688	-2.037	-0.545
	-0.352	-0.163	-0.342	-0.286	-0.244	-0.108	-0.251	-0.085
	(-0.971)	(-0.435)	(-0.943)	(-0.742)	(-0.897)	(-0.342)	(-0.899)	(-0.271)
Hisp	-2.235	-1.23	-2.658	-1.765	-5.676*	-1.303	-6.202*	-0.879
	-0.230	-0.159	-0.246	-0.117	-0.701*	-0.204	-0.764*	-0.137
	(-0.455)	(-0.304)	(-0.487)	(-0.222)	(-1.686)	(-0.513)	(-1.674)	(-0.367)
30K-60K	-2.235	-1.230	-2.658	-1.765	1.185	1.676	1.418	1.495
	-0.274	-0.194	-0.324	-0.277	0.146	0.262	0.175	0.234
	(-0.828)	(-0.486)	(-0.988)	(-0.704)	(0.518)	(0.709)	(0.597)	(0.627)
60K-100K	-5.045**	-3.727	-5.194**	-4.087*	3.183	2.837	3.311	2.994
	-0.618**	-0.587	-0.634**	-0.642*	0.393	0.444	0.408	0.468
	(-2.035)	(-1.614)	(-2.096)	(-1.771)	(1.444)	(1.330)	(1.500)	(1.422)
100K+	-3.854	-3.183	-4.028*	-3.476	0.197	0.263	0.161	0.76
	-0.473	-0.502	-0.491*	-0.546	0.024	0.041	0.020	0.119
	(-1.600)	(-1.365)	(-1.677)	(-1.564)	(0.099)	(0.134)	(0.080)	(0.394)
College	-0.599	-2.565	-0.826	-2.172	2.906	1.213	2.906	0.85
	-0.073	-0.404	-0.101	-0.341	0.359	0.19	0.358	0.133
	(-0.280)	(-1.332)	(-0.386)	(-1.147)	(1.284)	(0.548)	(1.157)	(0.396)
Postgrad	4.950*	2.559	4.81*	2.654	1.783	-0.055	2.138	-0.205
	0.607*	0.403	0.587*	0.417	0.220	-0.009	0.264	-0.032
	(1.724)	(0.977)	(1.672)	(1.045)	(0.853)	(-0.028)	(1.001)	(-0.104)
N	980	980	980	980	980	980	980	980
Pseudo R^2	0.043	0.034	0.043	0.036	0.033	0.027	0.034	0.028
Ref. Prob.	0.143	0.196	0.142	0.195	0.144	0.194	0.144	0.194

to (4), implying that the relationship between age and optimality of mutual fund choice is U-shaped. This result is not consistent with the finding by Lusardi and Mitchell (2011) that age has an inverted U-shape association with financial literacy. One possible explanation for this discrepancy is that mutual fund investors are not representative of the entire population in each age period. Mutual fund participation rates vary with age. ICI reports that age has an invert U-shaped effect on mutual fund ownership (Bogdan et al., 2010), being lowest for the young (18-34 years old) and the older (older than 65 years) groups, but peaks in the middle (35-64 years old) of the life cycle. Therefore, the young and older individuals who invest in mutual funds might possess more financial knowledge than their peers at their age.

When we turn our attention to the association between investor income and mutual fund choices, we find that the coefficients on *60K-100K* are negative and significant in columns (1), (3) and (4). The coefficients on *30K-60K* and *100K+* are also negative, consistently with the idea that high-income investors are more likely to avoid low-quality funds, although the coefficients are smaller than those on *60K-100K* and not significant. Therefore, the association between investor income and optimality of investor choices appears to be positive, consistently with the literature (Bailey et al., 2011; Engström, 2007), but not strictly monotonic. Tang et al. (2010) find that high-income investors experience smaller losses from inefficient investment choices.

Coefficients on ethnicity-based categories are all negative (*Caucasian* is the omitted variable), although only the coefficient on the African American category is statistically different from zero when the underperforming fund group consists of funds with predicted alpha in the bottom 20%. This result suggests that the worst funds are associated with lower fraction of African American investors and higher fraction of Caucasian investors. Moreover, we also report positive and significant coefficients of *Afram* in columns (5)-(8), which suggests that African Americans are strongly associated with top funds: A one percent shift in the fraction of African American investors increases the probability of a fund being good by 0.70%. Some previous studies show evidence that African Americans are associated with less financial literacy (See Bowen, 2008). However, according to Straight (2002), differences between African Americans and Caucasians are due to other socioeconomic characteristics, such as income or education. It should also be noted that African American mutual fund investors are not a representative sample of the whole African American population. African Americans both in our sample and in the mutual fund population are underrepresented (5% of mutual fund investors as opposed to 13.6% in the entire US population).[8] Therefore, we cannot easily extrapolate findings obtained from representative samples to the sample of

[8] http://www.census.gov/prod/cen2010/briefs/c2010br-06.pdf

African American mutual fund investors.

Finally, we find that education is not associated with better mutual fund choices. None of the coefficients are significant. If anything, there is a weakly positive association between the fraction of investors with postgraduate degrees and the probability of the fund belonging to the underperforming group. These results are consistent with those of Lusardi and Mitchell (2011), that education is not perfectly correlated with financial literacy. Grinblatt et al. (2011) find that, after controlling for fund families, there is no significant association between investor education and mutual fund fees.

Our investor characteristics data are aggregated at the management company level, i.e., funds in the same management company share the same investor characteristics. As a consequence, our proxy for investor characteristics is more representative of the characteristics of investors in funds with larger size and/or more shareholders within the company. However, in the regression analysis all funds receive the same weight. If overrepresented funds exhibit better performance than average, then both outperforming and underperforming funds in the fund family will be matched to the characteristics of sophisticated investors. On the other hand, if overrepresented funds exhibit worse performance than average, then both underperforming and outperforming funds will be matched to the characteristics of unsophisticated investors. Depending on how performance is distributed within management companies, aggregation could bias the relationship between investor characteristics and mutual fund performance in either direction. Of course, this problem would be less severe the fewer the number of funds per management company. As a robustness test, we repeat the analysis with a restricted sample that includes only families with fewer funds than the sample median. Table 3.8 shows the results. Although coefficients are not significant in some specifications, the results are generally consistent with those obtained for the full sample.

3.3.3 What Mistakes do Investors Make?

Our measure of predicted alpha is a function of four fund characteristics: size, age, expense ratio, and past performance. Therefore, if one investor characteristic is associated with poor predicted performance, that characteristic must necessarily be associated with one or more predictors of mutual fund performance, or with another variable that is correlated with predictors of performance. In order to gain further insight into the results of Table 3.7, we investigate why a higher fraction of certain sociodemographic groups characterize funds with poor predicted performance. More specifically, we ask which

Table 3.8: Robustness Test: Families with Few Funds
The table reports coefficients and marginal effects and associated z-values from a logit regression for the probability that the fund belongs to the group of funds with bottom/top predicted alpha. Regressions include only funds that belong to families whose number of funds in our sample is below the median. The first, second and third row of each variable reports coefficients, marginal effects and z-statistics, respectively. The dependent variable equals one if a fund's predicted alpha is at the bottom/top 15%/20%. Predicted alpha is calculated as the time-series mean of predicted alphas calculated by Fama-MacBeth (F-M) or pooled OLS (OLS) regressions from June, 2008 to June, 2010. All other variables are defined in Table 2. Standard errors are clustered by management company. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.

	Predicted Alpha			
	15% bottom (F-M)	20% bottom (F-M)	15% bottom (OLS)	20% bottom (OLS)
	(1)	(2)	(3)	(4)
Female	2.949	2.497*	3.524**	3.146**
	0.365	0.394*	0.430**	0.494**
	(1.585)	(1.724)	(2.170)	(2.081)
35-49	4.997**	5.666***	6.452***	5.862***
	0.618**	0.893***	0.787***	0.921***
	(2.148)	(2.927)	(2.824)	(2.954)
50+	3.303*	2.496*	3.216*	2.676**
	0.409*	0.393*	0.392*	0.420**
	(1.706)	(1.822)	(1.91)	(1.979)
Afram	-0.259	-5.134**	-3.863	-5.788**
	-0.032	-0.809**	-0.471	-0.909**
	(-0.129)	(-1.977)	(-1.414)	(-2.052)
Asian	-6.209**	-1.032	-2.807	-1.82
	-0.768**	-0.163	-0.342	-0.286
	(-2.452)	(-0.433)	(-0.926)	(-0.738)
Hisp	4.019	-1.009	-2.015	-0.743
	0.497	-0.159	-0.246	-0.117
	(0.789)	(-0.303)	(-0.486)	(-0.222)
30K-60K	-1.504	-1.23	-2.658	-1.765
	-0.186	-0.194	-0.324	-0.277
	(-0.553)	(-0.484)	(-0.984)	(-0.701)
60K-100K	-3.326	-3.727	-5.194**	-4.087*
	-0.412	-0.587	-0.634**	-0.642*
	(-1.384)	(-1.612)	(-2.135)	(-1.783)
100K+	-1.919	-3.183	-4.028*	-3.476
	-0.238	-0.502	-0.491*	-0.546
	(-0.786)	(-1.34)	(-1.67)	(-1.553)
College	0.762	-2.565	-0.826	-2.172
	0.094	-0.404	-0.101	-0.341
	(0.323)	(-1.356)	(-0.389)	(-1.162)
Postgrad	6.772**	2.559	4.81	2.654
	0.838**	0.403	0.587	0.417
	(2.114)	(0.961)	(1.608)	(1.033)
N	526	526	526	526
Pseudo R^2	0.057	0.034	0.043	0.036
Ref. Prob.	0.145	0.196	0.142	0.195

specific fund characteristics are associated with investor groups buying the worst performing funds. To answer this question, we estimate a logit model for the probability that the fund is in either tail of the distribution of each of the fund characteristics that predict fund performance. As explanatory variables we include the same investor characteristics as in Table 3.7, i.e., the percentage of investors of each sociodemographic group, as well as performance predictors other than the fund characteristic in question.

Table 3.9: Investor and Fund Characteristics
The table reports coefficients and marginal effects and associated z-values from a logit regression for the probability of the fund belonging to the group of funds with bottom/top fund characteristics corresponding to the first row of the table. The first, second and third row of each variable report coefficients, marginal effects and z-statistics, respectively. The dependent variable equals 1 if a fund's characteristic is at the top/bottom 15%. Marketing Fees equals 12b-1 fees plus one fourth of load fees. All other variables are defined in Tables 2 and 4. Standard errors are clustered by management company. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.

	Size		Exp Ratio		Age		Past Returns		Marketing	
	15% top	15% bottom	15% top	15% bottom	15% top	15% bottom	15% top	15% bottom	15% top	15% bottom
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	-2.935*	-1.592	2.435	-2.762*	-0.466	-0.651	-2.728*	3.061**	1.225	-8.980***
	-0.324*	-0.105	0.236	-0.284*	-0.065	-0.025	-0.320*	0.355**	0.106	-1.059***
	(-1.876)	(-0.635)	(1.428)	(-1.663)	(-0.341)	(-0.458)	(-1.794)	(2.115)	(0.624)	(-2.828)
35-49	6.391***	0.329	1.065	-3.584	-2.241	3.022	-1.445	6.460***	6.574*	-4.125
	0.706***	0.022	0.103	-0.369	-0.313	0.116	-0.170	0.750***	0.569*	-0.486
	(2.669)	(0.169)	(0.388)	(-1.541)	(-1.298)	(1.424)	(-0.900)	(3.047)	(1.881)	(-1.302)
50+	2.089	0.131	1.979	-1.647	-0.011	-0.205	-1.995	3.522**	5.123**	-5.782**
	0.231	0.009	0.192	-0.169	-0.001	-0.008	-0.234	0.409**	0.443**	-0.682**
	(1.191)	(0.088)	(1.019)	(-0.978)	(-0.009)	(-0.118)	(-1.627)	(2.332)	(2.229)	(-2.213)
Afram	-0.214	-1.452	-1.576	4.071*	-1.123	-1.262	5.215***	-6.234*	1.800	-7.749*
	-0.024	-0.096	-0.153	0.419*	-0.157	-0.049	0.612***	-0.724*	-0.156	-0.914*
	(-0.121)	(-0.615)	(-0.447)	(1.878)	(-0.417)	(-0.432)	(2.576)	(-1.953)	(-0.686)	(-1.652)
Asian	6.026***	-1.301	-4.708	2.932	-3.634	3.842*	-0.392	-3.289	-7.048**	1.590
	0.666***	-0.086	-0.456	0.301	-0.507	0.148*	-0.046	-0.382	-0.610**	0.187
	(2.605)	(-0.466)	(-1.484)	(1.513)	(-1.361)	(1.954)	(-0.185)	(-1.209)	(-2.053)	(0.439)
Hisp	1.431	-1.309	-5.457	2.047	2.204	2.419	-2.679	-0.178	7.069***	0.371
	0.158	-0.087	-0.529	0.210	0.307	0.093	-0.315	-0.021	0.612***	0.044
	(0.530)	(-0.256)	(-1.348)	(0.849)	(0.631)	(0.897)	(-0.919)	(-0.043)	(2.709)	(0.072)
30K-60K	1.935	-1.565	1.708	4.167	-1.317	-5.471*	2.546	-4.611	4.305	-0.407
	0.214	-0.104	0.166	0.429	-0.184	-0.211*	0.299	-0.535	0.373	-0.048
	(0.563)	(-0.587)	(0.636)	(1.289)	(-0.538)	(-1.954)	(0.966)	(-1.621)	(1.096)	(-0.086)
60K-100K	4.292	-7.460**	1.704	5.300*	-0.140	-6.881**	4.246*	-6.142**	5.782	2.865
	0.475	-0.494**	0.165	0.545*	-0.020	-0.265**	0.499*	-0.713**	0.501	0.338
	(1.418)	(-2.272)	(0.658)	(1.802)	(-0.051)	(-2.554)	(1.684)	(-2.428)	(1.627)	(0.678)
100K+	7.302**	-1.659	-1.753	5.911**	0.098	-2.718	3.679	-4.430**	6.442*	0.357
	0.807**	-0.110	-0.170	0.608**	0.014	-0.105	0.432	-0.514**	0.558*	0.042
	(2.442)	(-0.623)	(-0.740)	(2.190)	(0.050)	(-1.318)	(1.504)	(-2.046)	(1.859)	(0.095)
College	-4.188	-0.620	-0.366	-2.534	-0.987	-6.681**	-0.390	0.151	-5.375**	-0.099
	-0.463	-0.041	-0.036	-0.261	-0.138	-0.257**	-0.046	0.018	-0.465**	-0.012
	(-1.595)	(-0.192)	(-0.205)	(-0.891)	(-0.574)	(-2.279)	(-0.151)	(0.077)	(-1.964)	(-0.029)
Postgrad	-4.742**	-1.685	2.329	4.094	1.434	-7.044***	0.774	5.286***	-3.676	2.572
	-0.524**	-0.112	0.226	0.421	0.200	-0.271***	0.091	0.614***	-0.318	0.303
	(-2.123)	(-0.543)	(1.201)	(1.607)	(0.784)	(-2.739)	(0.356)	(2.736)	(-1.374)	(0.849)
Size			-0.569***	0.382***	0.494***	-0.422***	0.060	-0.152**	0.192**	-0.088
			-0.055***	0.039***	0.069***	-0.016***	0.007	-0.018**	0.017**	-0.010
			(-6.782)	(4.507)	(7.886)	(-4.600)	(0.593)	(-2.031)	(2.197)	(-1.008)
Exp Ratio	-1.763***	1.792***			0.081	-1.000*	0.889**	0.146	4.441***	-2.540***
	-0.195***	0.119***			0.011	-0.039*	0.104**	0.017	0.384***	-0.300***
	(-5.200)	(4.164)			(0.264)	(-1.847)	(2.167)	(0.463)	(5.952)	(-4.161)
Age	1.204***	-0.795***	-0.250*	-0.430*			-0.279	-0.138	-0.524**	0.199
	0.133***	-0.053***	-0.024*	-0.044*			-0.033	-0.016	-0.045**	0.023
	(7.730)	(-3.671)	(-1.720)	(-1.833)			(-1.626)	(-0.846)	(-2.084)	(1.290)
Past Returns	0.004	0.008**	-0.008	-0.013**	-0.003	0.010***			-0.020***	0.006
	0.000	0.001**	-0.001	-0.001**	-0.000	0.000***			-0.002***	0.001
	(1.180)	(2.011)	(-1.344)	(-2.008)	(-1.021)	(3.495)			(-2.783)	(1.059)
N	1061	1061	1061	1061	1061	1061	1061	1061	1061	1061
Pseudo-R^2	0.189	0.100	0.128	0.109	0.094	0.107	0.041	0.058	0.277	0.163
Ref. Prob.	0.127	0.071	0.108	0.116	0.168	0.040	0.136	0.134	0.096	0.137

Columns (1)-(8) of Table 3.9 report estimated coefficients and marginal effects from the logit regressions for the probability that the fund is in the bottom and top of the distribution of each one of the four performance predictors. The threshold to define the top and bottom of the distribution is 15%. The coefficients and marginal effects on *Female* are negative and significant for the probability that the fund has a low expense ratio and a high recent return, and are positive and highly significant for the probability that the fund has a low recent return. These findings suggest that the association between the fraction of female investors and poor predicted performance reported above happens both through female investors' presence in funds with poor recent performance, and through their absence from funds with low expenses and high recent performance. However, *Female* is also negatively associated with the probability that the fund is among the largest, which tend to have lower predicted performance.

In Table 3.7, we report a negative association between investor age, especially the *35-49* variable, and predicted performance. The results of Table 3.9 suggest that this association is driven by two forces that move investors in the same direction. First, a higher fraction of investors in the 35-49 age group is associated with a higher probability of the fund managing a large amount of assets. Second, investors in that age group, as well as older investors, are positively and significantly associated with poor recent performance.

Higher-income investors appear to be more present in funds with the lowest expense ratios and less present in funds with worst recent performance and the youngest funds. These results could explain the positive association between investor income and predicted performance. On the other hand, the fraction of higher-income investors (more clearly so, those in the 60K-100K income group) is negatively associated with the probability that the fund is among the smallest, which tend to perform better.

Table 3.9 also suggests that the percentage of African American investors in the fund is positively associated with low expense ratios and with good recent performance, consistently with the results of Table 3.7.

Finally, investors in the *Postgrad* group appear to avoid the largest and youngest funds, which tend to perform worse. However, they are drawn to funds with very poor recent returns, which explains why this group of investors is associated with poor predicted performance.

Taken together, these results indicate that it is mainly differences in sensitivity to past performance across investor sociodemographic groups and, in some cases, differences in sensitivity to expense ratios, what can explain systematic differences in the optimality of investor choices.

3.3.4 Fund marketing and investor choices

The results of columns (1)-(8) in Table 3.9 are useful to understand why the choices of some investor groups lead to worse predicted performance. However, they do not explain the reasons behind those choices, that is, why some investors fail to avoid funds that have underperformed in the past and can be predicted to underperform in the future. While a comprehensive answer to this important question is beyond the scope of this chapter, we may investigate whether management companies' actions play a role in investors' choices. More specifically, our data enable us to study whether there exists an association between marketing effort, as proxied by the fund's marketing fees, and the investment decisions of the disadvantaged clientele. Our question is motivated by the idea that the disadvantaged clientele is more likely to be influenced by management companies' marketing efforts, i.e., advertising and brokers' advice, which induce them to invest more than the rest of investors in underperforming funds. Indeed, a number of studies find an association between marketing and worse-quality funds. Bergstresser et al. (2009) report that funds distributed through brokers underperform their peers, even before distribution fees, in the direct distribution channel. Also, there is little evidence that mutual fund advertising signals future superior performance (Gallaher et al., 2006; Jain and Wu, 2000). A slightly different interpretation is that unsophisticated investors exhibit a stronger preference for financial advice, so fund families catering to these investors face a weaker incentive to generate alpha and, instead, spend resources on distribution (Del Guercio and Reuter, 2011). Under either of these hypotheses, we would expect funds with high marketing fees to attract the groups of investors that are associated with poor predicted performance.

An alternative explanation about the mechanism that drives the empirical results presented in this chapter is that the disadvantaged clientele lacks the knowledge, expertise, or skills to identify funds with poor past performance, while the rest of investors can identify those funds through their own research or through analysts' reports. Under this hypothesis, underperforming funds have a higher fraction of unsophisticated investors due to the sophisticated investors' decision to avoid those funds. In this case, we would expect no association between high marketing fees and the investor characteristics that are associated with poor predicted performance.

In columns (9) and (10) of Table 3.9, we estimate logit regressions for the probability that the fund's marketing fees are in the top and bottom 15% of the distribution, respectively. In both cases, we control for performance predictors. Following Sirri and Tufano (1998), we define annual marketing fees as 12b-1 fees plus the sum of front-end and back-end loads amortized over the average holding period.[9] The coefficient on *Female* is not significant for the probability that the fund has high marketing fees. These findings seem to favor the hypothesis that female investors' choices are not the consequence of this group of investor being more vulnerable to marketing.

The coefficients on *35-49* and *50+* for the probability that the fund charges top marketing fees are both positive and significant, which suggests that older investors' choice of recent underperforming funds could be explained by their preference for heavily marketed funds. This is especially the case for investors in the 50+ age group.

We do not find a negative association between the fraction of high-income investors and marketing effort, suggesting that the poorer choices of low-income investors are not explained by marketing. Finally, the coefficient on *Postgrad* for the probability that the fund charges high marketing fees is not positive, so the results for this group of investors reported above are not explained by marketing, either.

Therefore, heterogeneity in marketing sensitivity (or preference for marketed funds) across groups of investors cannot explain differences in the optimality of mutual fund choices found in this chapter, except for the case of older investors.

3.4 Conclusions

Although a number of previous studies assume the presence of unsophisticated investors in the mutual fund market, there is only limited evidence on the observable characteristics of those investors, mainly due to a lack of comprehensive databases on investor characteristics and their mutual fund choices. We overcome this limitation by using a novel data set of the sociodemographic characteristics of US Internet visitors to mutual fund websites to investigate the relation between mutual fund investor characteristics and investment decisions. We report evidence suggesting that mutual fund investors and visitors to mutual fund websites largely overlap, and consequently, the distribution of sociodemographic characteristics is similar in both populations.

[9] We amortize loads by dividing them by a holding period of four years because this is approximately the average holding period for US mutual fund investors in more recent years according to Bogle (2005).

Our chapter makes three contributions to the literature on investors' mutual fund choices. First, we report direct empirical evidence on the heterogeneity of investor sophistication in the US mutual fund market: Investor sociodemographic characteristics are systematically associated with funds' predicted performance. More specifically, funds with a higher fraction of female investors are more likely to have a low predicted performance. Also, older and low-income investors are associated with worse investment decisions. However, the relation between both age and income and the optimality of investor decisions is not strictly monotonic. We find evidence that the fraction African-American visitors is associated with top-performing funds. Interestingly, we do not find a positive relation between the optimality of mutual fund investment decisions and education levels.

Second, our results suggest that differences in mutual fund choices across investor sociodemographic groups arise in most cases from the fact that not all investors react in the same way to past performance and expense ratios.

Third, we find no clear evidence that differences in mutual fund choices across investor groups are driven by different sensitivities to fund marketing or different preferences for advice. Only the oldest investors exhibit a greater tendency to purchase heavily marketed funds.

Our findings are of interest to regulators in three ways. First, regulators may identify vulnerable investors *ex ante* on the basis of their sociodemographic characteristics, which are observable. This enables policy makers to target their policies to specific groups. Second, our analysis reveals the kind of information that is ignored by the "disadvantaged clientele" and, therefore, can be used as a guidance in the design of policies aimed at enhancing transparency. Finally, despite the evidence that marketing is associated with differences in net-of-fee fund performance across funds, our results do not suggest that differences in mutual fund choices across groups of investors are explained by differences in responsiveness to marketing effort.

4

Who are the Mutual Fund Dividend Clienteles?

Contents

4.1 Introduction

Since the publication of Miller and Modigliani's seminal paper, dividends have been one of the thorniest puzzles in corporate finance (Allen et al., 2000). One of the popular explanations of their presence is that some investors, particularly older and low income individuals, have a preference for dividends (Becker et al., 2011; Graham and Kumar, 2006). The literature terms those investors "dividend clienteles". However, despite the fact that the existence and the characteristics of dividend clienteles for stocks have been well-discussed, whether there are dividend clienteles for mutual funds and their characteristics are unresolved questions. The purpose of this chapter is to address those questions. More specifically, I use the sociodemographic characteristics of US visitors to mutual fund management companies' websites to proxy for investor characteristics and regress the dividend frequency and quantity on the investor characteristics.

Understanding who the dividend clienteles are in the mutual fund market is important for several reasons. First, investor behavior clearly affects mutual fund strategies, particularly, their investment styles and dividend policy. Desai and Jin (2011) document that the firm dividend policy is affected by the existence of dividend clienteles in the stock market. Similarly, if there are dividend clienteles in the mutual fund market, some mutual funds may try to attract those investors by deliberately choosing dividend policy. Moreover, dividends paid by firms vary across firm characteristics (such as size or age) that are associated with investment styles (Fama and French, 2001). To be able to implement a certain dividend policy, mutual funds have to adjust their investing styles. Second, the characteristics of dividend clienteles in the mutual fund market are helpful to identify unsophisticated investors. Harris et al. (2012) find evidence that mutual fund dividend quantity is associated with the presence of unsophisticated investors. As such, it is important for investor protection. Third, it is important for us to understand the segmentation of the mutual fund market. Limited competition and market segmentation are discussed in many previous studies.[1] In Chapter 2, I find evidence that if an investor's main goal is to obtain a regular income stream, he/she would care less about performance. Therefore, it is possible that mutual funds segment the market by their dividend policy. This may provide an explanation of why some investors do not leave the funds with worst performance.

Previous literature documents the existence of dividend clienteles in the stock

[1] See, for example, Huang et al. (2007), Gil-Bazo and Ruiz-Verdú (2009) and Del Guercio et al. (2010).

market by showing that certain types of investors have a preference for dividends. Miller and Modigliani (1961) and Shefrin and Thaler (1988) are the first to conjecture that retired (older) investors, or any other investors with a pronounced need for a regular income stream, would prefer income stocks for consumption purposes. These authors argue that, because of self-control, dividend clienteles may prefer dividends over "home made" dividends, i.e., income generated by the partial liquidation of the holdings. Consistent with their arguments and conjectures, Graham and Kumar (2006) and Becker et al. (2011) provide strong empirical evidence of the existence of dividend clienteles by showing that older and low income investors are likely to favor stocks with dividend distributions.

However, the mutual fund market is different from the stock market in several aspects. First of all, the composition of investors is different between two markets. The Investment Company Institute (ICI) reports that an estimated 92 million individual investors owned mutual funds in the US at the end of 2012. And those investors held 89 percent of total mutual fund assets. However, less than 30 percent of total stock assets are held by individual investors.[2] The dominating position of institutional investors in the stock market suggests that corporate dividend policy is used to cater to institutional investors rather than retail investors. Second, mutual fund dividend payout is, as described in Chapter 2, highly regulated while a corporate has more flexibility to pay dividends. In the US, mutual funds are registered as regulated investment companies. The mutual funds do not need to pay certain types of corporate taxes if they fulfill some restrictions. One of the restrictions is that a mutual fund has to pay out all the dividends collected from stocks at the end of the year. Therefore, the differences in the characteristics of dividend payers are large. For example, Denis and Osobov (2008) document that in the stock market, dividends are concentrated among the largest, most profitable companies in the stock market: Less than 20% of stocks pay dividends in 2002. Yet, paying dividends seems to be prevalent among mutual funds because of the pass-through regulation. In my sample, more than 63% of funds pay dividends in a single year. Therefore, whether dividend clienteles for mutual funds exist and the composition of these clienteles are open to questions.

In Chapter 2, I presume that dividend clienteles exist and their motivation is to generate a regular income stream to finance consumption. I test this joint hypothesis and some empirical evidence supporting it. I find that the sensitivity of flows to participation cost varies across the dividend policy, i.e., a multi-dividend fund has a smaller outflow than a single-dividend funds given participation cost. Harris et al. (2012) document similar results in terms of dividend quantity. They show that some funds purchase stocks before divi-

[2] http://www.learningmarkets.com/understanding-institutional-ownership/

dend payment dates to artificially increase their dividend yields. However, the authors argue that the theory of dividend clientele cannot explain this anomaly because it is not the optimal way to collect the dividends. Their argument is based on the assumption that the dividend clienteles should only collect dividends without paying any costs. However, by definition, dividend clienteles are willing to trade some costs for dividends. For example, a clientele always needs to pay the tax if he/she receives dividends from a stock even if he/she has a low tax rate. The results documented by Harris et al. (2012), to the contrary, support the theory of dividend clientele. To further investigate the dividend clienteles in the mutual fund market, I try to provide direct evidence by documenting who those dividend clienteles are, i.e., the sociodemographic characteristics of dividend clienteles.

This paper also contribute to literature by examining the motivation of dividend clienteles. Previous studies suspect that investors have two motivations to prefer dividends: Investors might use dividends to finance their consumption and/or they have a lower tax rate.[3] However, there is no direct evidence of what really motivates dividend clienteles. The literature shows the existence of dividend clienteles by documenting that older and low income investors are more likely to be dividend clienteles in the stock market. Yet those investors who typically have pronounced needs to finance their consumption also have lower tax rates at the same time. As Becker et al. (2011) state in their paper, "...older investors likely have more pronounced needs to 'cash out' larger portions of their portfolios for consumption and also likely have lower tax rates on dividends relative to younger investors... In this paper, we do not address the reasons for seniors' preference for dividends...". The existence of dividend clienteles in the mutual fund market helps to examine their motivation. Because of the special regulation of mutual fund dividend distribution (e.g., pass through regulation), mutual fund dividend distribution has two dimensions: frequency and quantity. Those two attributes can be interpreted as corresponding to the two motivations for a dividend preferences described above. If a dividend clientele invests in a mutual fund for the purpose of obtaining a regular income stream, he/she would choose the mutual funds with more frequent dividend distributions, but not the funds with large dividend distributions. The dividend clientele could obtain the amount of dividend as he/she wants by changing the initial investment. For example, if an investor expect to receive \$100 per year, he/she can either invest in 10 shares of a fund with \$10 dividend or 100 shares of another fund with \$1 dividend. Yet the dividend clientele cannot decide the effect of dividend frequency if we implicitly assume that dividend clienteles prefer receiving dividends over partially selling funds as Shefrin and Thaler (1988) argue. On the other side, if a dividend clien-

[3] See, for example, Shefrin and Thaler (1988), Graham and Kumar (2006) and Becker et al. (2011) for more details.

tele prefer dividends because he/she has a low tax rate, the dividend clientele might not be concerned about dividend frequency because the dividend frequency has no effect or, at most, a small effect, on the taxes he/she pays. Therefore, the dividend clientele may prefer the funds with large dividend distributions. The results in Chapter 2 show some evidence of the consumption rationale. This chapter provides more empirical evidence of the motivation of dividend clienteles in the mutual fund market.

The main results of this chapter are as follows. First, I document some evidence of the existence of dividend clienteles in the mutual fund market. I find that low income and older investors are associated with funds paying dividend more frequently. This result is consistent with my expectation and previous studies. Second, I do not find any significant evidence of a relationship between the size of dividend distributions and investor characteristics. This result plausibly suggests that the tax rationale cannot explain that existence of dividend clienteles in the mutual fund market.

4.2 Data

I use the sociodemographic characteristics of US visitors to mutual fund websites to proxy for investor characteristics. I obtain the the sociodemographic characteristics of US visitors to mutual fund websites from Quantcast, an Internet audience measurer. Quantcast provides provides information on the sociodemographic composition of visitors to Internet websites. For example, 62% of the visitors to the website of Vanguard asset management company are male while the other 38% are female. Quantcast estimates the sociodemographic characteristics of visitors to websites using two sources of databases. The first one is the directly measured data, which covers 10 million website destinations and over 200 million people in the US. The second one is multiple reference data which covers over 1.5 millions individuals in the US. Quantcast has non-personally identifiable information user data and the information on their click-stream. Quantcast uses statistical methods to combine the two databases. For more detail, please check the Quantcast Methodology Overview[4].

I first use public information to obtain the websites for mutual fund management companies. I input the websites into Quantcast to collect the characteristics of the visitors to each website. Because there are few visitors to the websites of mutual fund management companies that are extremely small,

[4] http://www.quantcast.com/white-papers/quantcast-methodology.pdf

I only keep the largest 250 mutual fund management companies in terms of assets under management. Finally, I obtain sociodemographic characteristics of visitors to websites belonging to 207 management companies as of June 2010. Chapter 3 provides more details and summary statistics. Table 3.3 provides evidence that visitors to mutual fund websites and mutual fund investors largely overlap. The sociodemographic characteristics of visitors to mutual fund websites are close to the characteristics of mutual fund investors but far from the characteristics of Internet users. This evidence suggests that the sociodemographic characteristics of US visitors to mutual fund websites are a good proxy for investor characteristics.

I obtain the mutual fund data from CRSP Survivor-Bias Free Mutual Fund Database. Because the investor characteristics are obtained in 2010, I perform the fund characteristics as cross-sectional data in 2010. Following standard use in the mutual fund literature, I focus on cap-based and style-based domestic equity mutual funds defined by the Lipper objective codes, i.e., LCVE, MLVE, EI, EIEI, LCCE, MLCE, LCGE, MLGE, MCVE, MCCE, MCGE, SCVE, SCCE, and SCGE. I also drop the institutional and index funds following CRSP identifiers, i.e., *index_fund_flag* and *inst_fund*. A mutual fund may be counted multiple times in the database if it has several share classes. I eliminate such redundant observations by computing the fund characteristics as the asset-weighted means of class characteristics if the classes share the same fund identifier code, *crsp_cl_grp*, provided by CRSP. Following Elton et al. (2011) and Evans (2010), I drop the funds with total net assets below $15 million and whose age is less than 36 months. Finally, the database contains a sample of 2,153 distinct funds.

Table 4.1 provides summary statistics. I identify dividend distributions as mutual fund dividends if the first letter of *dis_type* in CRSP is “D”. I define dividend ratio as the distribution amount over reinvestment price and the variable DivRatio as the yearly dividend ratio, which is calculated as the sum of dividend ratio during a calendar year. LnDivRatio is the natural logarithm of the dividend ratio plus 1. Some mutual funds pay dividends several times in a short period or even in the same day, typically when dividends belong to different types (e.g., income dividend and qualified income dividend). Therefore, I consider dividend distributions in the same month as one time. I count the numbers of months when a mutual fund pays dividends (from 0 to 12) as DivFreq. I define a multi-dividend dummy, MultiDiv, which is equal to 1 if dividend frequency is larger than one, and zero otherwise. The definitions of other variables of interest are the same as in the previous literature. LnSize is the nature logarithm of total net asset under fund management. LnAge is the nature logarithm of age in months. FrontLoad is the front load. ExpRatio is the expense ratio defined as total operating expenses divided by the year-end

Table 4.1: Summary Statistics
This table presents the summary statistics for our sample. DivRatio is the sum of the dividend ratio, defined as distribution amount over reinvestment price, for each dividend distribution during the whole year. LnDivRatio is the natural logarithm of the dividend ratio plus 1. DivFreq is the number of months in which a fund pays dividends. MultiDiv is a dummy variable that equals 1 if DivFreq is larger than 1, and 0 if DivFreq is equal to 1. LnSize is the nature logarithm of total net asset under fund management. LnAge is the natural logarithm of age in months. FrontLoad is the front load. ExpRatio is the expense ratio defined as total operating expenses divided by the year-end total net assets. TurnRatio is the turnover ratio. FundRet is the raw return for the past 12 months (buy and hold).

Summary Statistics						
	Mean	SD	Median	1st perc.	99th perc.	N
DivRatio (%)	0.533	0.865	0.189	0.000	3.273	2153
LnDivRatio (%)	0.528	0.839	0.189	0.000	3.221	2153
DivFreq (times)	1.643	1.286	1	1	5	2153
MultiDiv	0.367	0.482	0	0	1	2153
LnSize ($ million)	5.844	1.542	5.751	3.049	9.920	2096
LnAge (month)	5.025	0.583	5.017	3.932	6.756	2153
FrontLoad (%)	1.459	2.127	0.000	0.000	5.750	2108
ExpRatio (%)	1.251	0.394	1.279	0.190	2.260	1476
TurnRatio (%)	80.136	91.255	59.000	3.000	427.000	1464
Fundret (%)	18.677	6.362	17.149	8.008	34.292	1761

total net assets. TurnRatio is the turnover ratio. FundRet is the raw return for the past 12 months (buy and hold). To fully understand the database, I provide the correlation coefficients among variables of interest. Table 4.2 shows the results. Female is the percentage of female visitors. 35-49 is the percentage of visitors from 35 to 49 years old. 50+ is the percentage of visitors above 50 years old. Afram is the percentage of African American visitors. Asian is the percentage of Asian visitors. Hisp is the percentage of Hispanic visitors. 30K-60K is the percentage of visitors with income of 30-60 thousand dollars per year. 60K-100K is the percentage of visitors with income of 60-100 thousand dollars per year. 100K+ is the percentage of visitors with income of more than 100 thousand dollars per year. College is the percentage of visitors with college degree. Postgrad is the percentage of visitors with postgraduate degree. It is worth noting that *LnDivRatio* is positively associated with *MultiDiv*. This correlation is consistent with my argument in Chapter 2: Mutual funds with high undistributed dividends are more likely to use multi-dividend policy.

Table 4.2: Correlation Table
The table reports correlation coefficients among variables of interest. LnDivRatio is the natural logarithm of yearly dividend ratio plus one where yearly dividend ratio is defined as the sum of the dividend ratio, i.e., distribution amount over reinvestment price. DivFreq is the number of months in which a fund pays dividends. MultiDiv is a dummy variable that equals 1 if DivFreq is larger than 1, and 0 if DivFreq is equal to 1. Female is the percentage of female visitors. 35-49 is the percentage of visitors from 35 to 49 years old. 50+ is the percentage of visitors above 50 years old. Afram is the percentage of African American visitors. Asian is the percentage of Asian visitors. Hisp is the percentage of Hispanic visitors. 30K-60K is the percentage of visitors with income of 30-60 thousand dollars per year. 60K-100K is the percentage of visitors with income of 60-100 thousand dollars per year. 100K+ is the percentage of visitors with income of more than 100 thousand dollars per year. College is the percentage of visitors with college degree. Postgrad is the percentage of visitors with postgraduate degree.

	LnDivRatio	MultiDiv	Female	35-49	50+	Afram	Asian	Hisp	30K-60K	60K-100K	100K+	College	Postgrad
LnDivRatio	1												
MultiDiv	0.4028	1											
Female	-0.05	0.0279	1										
35-49	0.034	-0.0211	-0.0369	1									
50+	0.038	0.0088	-0.1614	-0.572	1								
Afram	0.0235	0.0357	-0.0043	0.1921	-0.2351	1							
Asian	0.0299	0.0118	0.1122	0.0384	-0.0004	-0.1602	1						
Hisp	-0.0388	-0.043	0.0441	0.1263	-0.2965	0.2024	-0.0389	1					
30K-60K	-0.0595	-0.011	0.0507	-0.0379	-0.1568	0.1552	-0.0846	0.1984	1				
60K-100K	-0.0216	-0.0276	0.1603	0.0576	0.0258	-0.2619	-0.0859	-0.2284	-0.4894	1			
100K+	0.0981	0.006	-0.2409	0.0309	0.3285	-0.1313	0.1524	-0.1291	-0.4966	-0.2398	1		
College	-0.0001	-0.0229	-0.0894	0.1418	-0.248	-0.0839	0.0932	0.1778	0.0994	-0.1183	0.0144	1	
Postgrad	0.0767	-0.0159	-0.3584	0.075	0.1062	-0.0384	0.2034	-0.0683	-0.1347	-0.234	0.4298	-0.2888	1

4.3 Empirical Strategy

The main purpose of this chapter is to identify the sociodemographic characteristics of mutual fund investors who exhibit a preference for dividends. Previous studies on dividend clienteles in the stock market identify the characteristics of investors who prefer dividends, measured in terms of dividend quantity. Yet, because of the pass-through dividend policy, mutual fund dividend payout has two dimensions: frequency and quantity. Those dimensions are correlated but different. Admittedly, the dividend frequency of a mutual fund is positively associated with dividend quantity as I show in Table 4.2. However, funds with high undistributed dividends need not distribute them several times during the year. They can always keep the dividends and pay them out at the end of the year.[5] At the same time, it is possible that an investor might prefer funds with more frequent dividend payouts but not with a large dividend payout, or vice versa. In this chapter, I will consider both types of investors, those who prefer more frequent dividend distributions, and those who prefer a huge amount of dividends paid during the year. I use both dividend frequency and dividend quantity as dependent variables.

I estimate a logit model to investigate how the sociodemographic characteristics of investors are associated with the frequency of fund dividend payouts. More specifically, I model the probability of a fund using multi-dividend policy as a function of the characteristics of the investors in the fund's family, controlling for fund characteristics. To model the association between dividend quantity and investor characteristics, I regress the quantity of dividends paid in 2010 on investor characteristics.

4.4 Empirical Results

Table 4.3 reports the results of the regressions of fund dividend payout, (e.g., frequency or quantity), on investor characteristics. Column (1) and (2) show the coefficients, marginal effects, and associated z-values from a logit regression for the probability that the fund follows a multi-dividend policy. The dependent variable is *MultiDiv*. Column (3) and (4) report the coefficients and t-statistics of OLS regression of dividend quantity, proxied by *LnDivRatio*, on investor characteristics. *LnDivRatio* is the natural logarithm of the sum

[5] See the discussion in Chapter 2.

of the dividend ratio for each dividend distribution during the whole year plus one where dividend ratio is defined as distribution amount over reinvestment price.

The coefficients for *35-49* and *50+* are positive and significant for *50+* when the dividend frequency is the proxy of dividend clienteles in columns (1) and (2). This results suggest that older investors are positively associated with dummy for dividend frequency. Or in other words, older investors are more likely to hold multi-dividend funds. A one percent increase in the proportion of investors in the *50+* age groups increases the probabilities of a fund using multi-dividend policy by 1.2%.

We find that the coefficients for income group varialbes *30K-60K*, *60K-100K* and *100K+* are negative in columns (1) and (2), though the coefficients on *60K-100K* are not significant. A one percent shift in the fraction of *30K-60K*, *60K-100K* and *100K+* investors decrease the probability of a fund using multi-dividend policy by 1.7%, 1.3% and 2.1%, respectively. The results are consistent with the idea that low-income investors are more likely to hold multi-dividend funds.

In sum, I find that the results are partly consistent with my expectation and previous literature (Becker et al., 2011; Graham and Kumar, 2006), that older and low-income investors are likely to be dividend clienteles, measured in terms of dividend frequency.

Interestingly, no coefficient is statistically significant at conventional significance levels if dividend quantity is the dependent variable, as reported in columns (3) and (4) of table 4.3 when the proxy for dividend clienteles is the dividend quantity. This result cannot be explained by the tax rationale theory. If an investor prefers dividends because he/she has a lower tax rate, he/she should be sensitive to quantity rather than dividend frequency. However, if dividend clienteles buy mutual funds for the purpose of receiving regular dividend streams, they would, as I argue above, care more about dividend frequency. This is consistent with my argument in Chapter 2.

4.5 Conclusions and Future Research

This chapter identifies the characteristics of dividend clienteles in the mutual fund market by using the sociodemographic characteristics of US visitors to mutual fund websites to proxy for investor characteristics. The results suggest that low income and older investors are associated with multi-dividend policy

Table 4.3: Investor Characteristics and Dividend Clienteles
The table reports the results of the regressions of fund dividend payout (measured by the frequency or quantity of dividend distributions) on investor characteristics. Columns (1) and (2) report the coefficients, marginal effects, and associated z-values from a logit regression for the probability that the fund uses multi-dividend policy. The first, second and third row of each variable report coefficients, marginal effects and z-statistics, respectively. The dependent value is *MultiDiv* defined as dummy variable that equals one if dividend frequency is larger than 1, and 0 otherwise. Columns (3) and (4) report the coefficients and t-statistics of OLS regressions of *LnDivRatio* on investor characteristics. *LnDivRatio* is the natural logarithm of the sum of the dividend ratio for each dividend distribution during the whole year plus 1 where dividend ratio is defined as distribution amount over reinvestment price. Female is the percentage of female visitors. 35-49 is the percentage of visitors from 35 to 49 years of age. 50+ is the percentage of visitors above 50 years of age. Afram is the percentage of African American visitors. Asian is the percentage of Asian visitors. Hisp is the percentage of Hispanic visitors. 30K-60K is the percentage of visitors with income of 30-60 thousand dollars per year. 60K-100K is the percentage of visitors with income of 60-100 thousand dollars per year. 100K+ is the percentage of visitors with income of more than 100 thousand dollars per year. College is the percentage of visitors with college degree. Postgrad is the percentage of visitors with postgraduate degree. The fund characteristics controls include the stability of the dividend payout, expense ratio, yearly relative flows, size, turnover ratio and fund raw performance defined in Chapter 3. Standard errors are clustered by management company. *** indicates significance at 1% level, ** indicates significance at 5% level and * indicates significance at 10% level.

	MultiDiv (1)	MultiDiv (2)	LnDivRatio (3)	LnDivRatio (5)
Female	2.333	3.305	-0.007	-0.008
	0.583	0.826		
	(0.928)	(1.299)	(-0.996)	(-1.182)
35-49	2.134	2.349	-0.011	-0.012
	0.533	0.587		
	(0.723)	(0.802)	(-1.010)	(-1.113)
50+	4.517**	4.794**	-0.000	-0.003
	1.129**	1.199**		
	(2.188)	(2.280)	(-0.078)	(-0.518)
Afram	3.412	2.682	0.037	0.035
	0.853	0.670		
	(1.206)	(1.075)	(1.331)	(1.341)
Asian	-2.659	-3.777	0.014	0.015
	-0.665	-0.944		
	(-0.859)	(-1.238)	(1.354)	(1.505)
Hisp	0.330	1.181	-0.013	-0.013
	0.082	0.295		
	(0.106)	(0.364)	(-1.457)	(-1.550)
30K-60K	-6.824*	-6.950*	0.009	0.013
	-1.705*	-1.737*		
	(-1.667)	(-1.673)	(0.525)	(0.704)
60K-100K	-5.284	-5.640	0.013	0.016
	-1.320	-1.410		
	(-1.259)	(-1.347)	(0.903)	(1.088)
100K+	-8.569**	-8.833**	0.009	0.013
	-2.141**	-2.208**		
	(-2.164)	(-2.181)	(0.618)	(0.881)
College	0.016	0.446	-0.005	-0.005
	0.004	0.111		
	(0.005)	(0.149)	(-0.710)	(-0.676)
Gradsch	-0.186	-0.217	0.002	0.002
	-0.046	-0.054		
	(-0.058)	(-0.065)	(0.338)	(0.344)
LnDivRatio		93.692***		
		23.422***		
		(4.054)		
MultiDiv				0.003**
				(2.545)
Fund Characteristics	Yes	Yes	Yes	Yes
Segment Fix Effect	Yes	Yes	Yes	Yes
N	506	506	513	513
Pseudo R^2	0.203	0.231	0.246	0.263

but not with a large quantity of dividends. A plausible explanation for these results is that dividend clienteles in the mutual fund market exist because they use dividends to finance their consumption, not because they have low tax rates.

Fama and French (2001) document the phenomenon that corporates pay less and less dividends over time. "The proportion of corporates paying cash dividends falls from 66.5% in 1978 to 20.8% in 1999" while Desai and Jin (2011) document a similar number, 19.1% in 2002. However, the proportion of mutual funds paying dividends increases from 46.2% in 2000 to 67.7% in 2011. The sharp contrast suggests that the demand for dividends in the stock market decreases while it increases in the mutual fund market. It is interesting to wonder why. One possible explanation is that the demand for dividends mainly comes from individual investors. When individual investors leave the stock market and enter the mutual fund market, the demand for dividends is transferred from the stock market to the mutual fund market. This conjecture may be helpful in explaining the long-time unsolved dividend puzzle. More specifically, the phenomenon cannot be explained by signalling motives or catering theory (Baker and Wurgler, 2004) but only the dividend clienteles.[6]

Another interesting topic is associated with investors psychological sentiment. Mutual funds pay two types of incomes to investors: dividend and capital gain. Unlike stocks, mutual fund investors receive capital gains passively, i.e., investors do not need to actively sell mutual funds, whereas stock investors receive capital gains only when they actively realize the appreciation of their holdings. It is still not clear whether mutual fund investors distinguish dividends from capital gains as proposed by Shefrin and Statman (1984). If mutual fund investors could recognize the differences between dividends and capital gains, another interesting question arises: Do mutual fund dividend clienteles prefer capital gains over dividends? Capital gains may be a better alternative to dividends to generate regular income streams because the tax rate on capital gains is always lower than that on dividends. Therefore, it appears as if, a clientele would wish to receive capital gains rather than dividends.

[6]DeAngelo et al. (2008) provide a survey on possible explanations on the dividend puzzle.

Bibliography

Alexander, G., J. Jones, and P. Nigro (1998). Mutual fund shareholders: Characteristics, investor knowledge, and sources of information. *Financial Services Review 7*(4), 301–316.

Allen, F., A. E. Bernardo, and I. Welch (2000). A theory of dividends based on tax clienteles. *The Journal of Finance 55*(6), 2499–2536.

Bailey, W., A. Kumar, and D. Ng (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics 102*(1), 1.

Baker, M. and J. Wurgler (2004). A catering theory of dividends. *The Journal of Finance 59*(3), 1125–1165.

Barber, B., T. Odean, and L. Zheng (2005). Out of sight, out of mind: The effects of expenses on mutual fund flows. *The Journal of Business 78*(6), 2095–2120.

Barclay, M., N. Pearson, and M. Weisbach (1998). Open-end mutual funds and capital-gains taxes. *Journal of Financial Economics 49*(1), 3–43.

Bauer, R. and P. Smeets (2010). Some men invest like women: The influence of social values on investment decisions and investor loyalty. *Working paper*.

Becker, B., Z. Ivković, and S. Weisbenner (2011). Local dividend clienteles. *The Journal of Finance 66*(2), 655–683.

Bergstresser, D., J. Chalmers, and P. Tufano (2009). Assessing the costs and benefits of brokers in the mutual fund industry. *Review of Financial Studies 22*(10), 4129.

Bergstresser, D. and J. Poterba (2002). Do after-tax returns affect mutual fund inflows? *Journal of Financial Economics 63*(3), 381–414.

Bergstresser, D., J. Poterba, and R. Zarutskie (2003). Mutual fund portfolio turnover and the effective tax burden on taxable investors. Working paper, Duke University.

Berk, J. and I. Tonks (2007). Return persistence and fund flows in the worst performing mutual funds. Working paper.

Bogdan, M., J. Sabelhaus, and D. Schrass (2010, Septermber). Ownership of mutual funds, shareholder sentiment, and use of the internet, 2010. *Investment Company Institute Fundamentals 19*(6), available at http://www.ici.org/pdf/fm–v19n6.pdf.

Bogle, J. (2005). The mutual fund industry 60 years later: For better or worse? *Financial Analysts Journal*, 15–24.

Bover, O. (2010). Wealth inequality and household structure: US vs. Spain. *Review of Income and Wealth 56*(2), 259–290.

Bover, O., J. M. Casado, S. Costa, P. Du Caju, Y. McCarthy, E. Sierminska, P. Tzamourani, E. Villanueva, and T. Zavadil (2013). The distribution of debt across euro area countries: The role of individual characteristics, institutions and credit conditions. Working paper.

Bowen, C. F. (2008). Money matters of african americans. In *Handbook of consumer finance research*, pp. 287–302. Springer.

Brown, K., W. Harlow, and L. Starks (1996). Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance*, 85–110.

Busse, J. A. (2001). Another look at mutual fund tournaments. *Journal of financial and Quantitative Analysis 36*(1), 53–74.

Campbell, J. (2006). Household finance. *The Journal of Finance 61*(4), 1553–1604.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), pp. 57–82.

Chen, J., H. Hong, M. Huang, and J. D. Kubik (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *The American Economic Review 94*(5), 1276–1302.

Chevalier, J. and G. Ellison (1997). Risk taking by mutual funds as a response to incentives. *The Journal of Political Economy 105*(6), 1167–1200.

Choi, J. J., D. Laibson, and B. C. Madrian (2010). Why does the law of one price fail? an experiment on index mutual funds. *Review of Financial Studies 23*(4), 1405–1432.

Christoffersen, S., C. Geczy, D. Musto, and A. Reed (2005). Crossborder dividend taxation and the preferences of taxable and nontaxable investors: Evidence from canada. *Journal of Financial Economics 78*(1), 121–144.

Christoffersen, S. and D. Musto (2002). Demand curves and the pricing of money management. *The review of financial studies 15*(5), 1499–1524.

Christoffersen, S. E., R. Evans, and D. K. Musto (2013). What do consumers' fund flows maximize? evidence from their brokers' incentives. *The Journal of Finance 68*(1), 201–235.

Cremers, M., M. Ferreira, P. Matos, and L. Starks (2011). The mutual fund industry worldwide: Explicit and closet indexing, fees, and performance. Working paper, Social Science Research Network.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *The Journal of Finance 66*(5), 1461–1499.

DeAngelo, H. and L. DeAngelo (2006). The irrelevance of the mm dividend irrelevance theorem. *Journal of Financial Economics 79*(2), 293 – 315.

DeAngelo, H., L. DeAngelo, and D. Skinner (2008). Corporate payout policy. In *Foundations and trends in finance*, Volume 3.

Del Guercio, D. and J. Reuter (2011). Mutual fund performance and the incentive to generate alpha. *Journal of Finance, forthcoming.*

Del Guercio, D., J. Reuter, and P. A. Tkac (2010). Broker incentives and mutual fund market segmentation. Working paper.

Denis, D. J. and I. Osobov (2008). Why do firms pay dividends? International evidence on the determinants of dividend policy. *Journal of Financial Economics 89*(1), 62–82.

Desai, M. A. and L. Jin (2011). Institutional tax clienteles and payout policy. *Journal of Financial Economics 100*(1), 68–84.

Elton, E., M. Gruber, and C. Blake (2011). Does size matter? the relationship between size and performance. Working paper, Fordham University Schools of Business.

Engström, S. (2007). Preferences and characteristics of mutual fund investors. Working paper, mimeo.

Engström, S. and A. Westerberg (2004). Information costs and mutual fund flows. Technical report, SSE/EFI Working paper series in Economics and Finance.

Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance 65*(4), 1581–1611.

Fama, E. F. and K. R. French (2001). Disappearing dividends: changing firm characteristics or lower propensity to pay? *Journal of Financial economics 60*(1), 3–43.

Fama, E. F. and K. R. French (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance 65*(5), 1915–1947.

Ferreira, M. A., A. Keswani, A. F. Miguel, and S. B. Ramos (2012). The flow-performance relationship around the world. *Journal of Banking & Finance 36*(6), 1759 – 1780.

Gallaher, S., R. Kaniel, and L. Starks (2006). Madison avenue meets wall street: Mutual fund families, competition and advertising. Woking paper, Duke University.

Gil-Bazo, J. and P. Ruiz-Verdú (2008). When cheaper is better: Fee determination in the market for equity mutual funds. *Journal of Economic Behavior & Organization 67*(3), 871–885.

Gil-Bazo, J. and P. Ruiz-Verdú (2009). The relation between price and performance in the mutual fund industry. *The Journal of Finance 64*(5), 2153–2183.

Graham, J. and A. Kumar (2006). Do dividend clienteles exist? evidence on dividend preferences of retail investors. *The Journal of Finance 61*(3), 1305–1336.

Grinblatt, M., S. Ikäheimo, M. Keloharju, and S. Knüpfer (2011). IQ and Mutual Fund Choice. Working paper, UCLA.

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance 51*(3), 783–810.

Harris, L., S. Hartzmark, and D. Solomon (2012). Juicing the dividend yield: Mutual funds and the demand for dividends. Working paper, University of Southern California.

Hartzmark, S. M. and D. H. Solomon (2013). The dividend month premium. *Journal of Financial Economics 109*(3), 640–660.

Hastings, J. S. and L. Tejeda-Ashton (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from Mexico. Working paper, National Bureau of Economic Research.

Huang, J., K. D. Wei, and H. Yan (2007). Participation costs and the sensitivity of fund flows to past performance. *The Journal of Finance 62*(3), 1273–1311.

Jain, P. and J. Wu (2000). Truth in mutual fund advertising: Evidence on future performance and fund flows. *The Journal of Finance 55*(2), 937–958.

Johnson, W. T. (2004). Predictable investment horizons and wealth transfers among mutual fund shareholders. *The Journal of Finance 59*(5), 1979–2012.

Johnson, W. T. and J. M. Poterba (2010). The effect of taxes on shareholder inflows around mutual fund distribution dates. Working paper, National Bureau of Economic Research.

Kumar, A., A. Niessen-Ruenzi, and O. Spalt (2012). What is in a name? mutual fund flows when managers have foreign sounding names. Working paper, SSRN.

Lusardi, A. and O. Mitchell (2011). Financial literacy around the world: an overview. Working paper, National Bureau of Economic Research.

Miller, M. and F. Modigliani (1961). Dividend policy, growth, and the valuation of shares. *Journal of business*, 411–433.

Niebling, F., S. Meyer, and A. Hackethal (2009). Whose money is smart? mutual fund purchases of private investors. Working paper, SSRN.

Scholz, J. K. (1992). A direct examination of the dividend clientele hypothesis. *Journal of Public Economics 49*(3), 261–285.

Shefrin, H. M. and M. Statman (1984). Explaining investor preference for cash dividends. *Journal of Financial Economics 13*(2), 253–282.

Shefrin, H. M. and R. H. Thaler (1988). The behavioral life-cycle hypothesis. *Economic inquiry 26*(4), 609–643.

Shiller, R. J. (2006). *Irrational exuberance*. Princeton University Press.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *The Journal of Finance 53*(5), pp. 1589–1622.

Straight, R. (2002). Wealth: Asset-accumulation differences by race: Scf data, 1995 and 1998. *The American Economic Review 92*(2), 330–334.

Tang, N., O. S. Mitchell, G. R. Mottola, and S. P. Utkus (2010). The efficiency of sponsor and participant portfolio choices in 401(k) plans. *Journal of Public Economics 94*(11), 1073–1085.

Thaler, R. (1999). The end of behavioral finance. *Financial Analysts Journal*, 12–17.

Thaler, R. H. and H. M. Shefrin (1981). An economic theory of self-control. *Journal of Political Economy 89*(2), pp. 392–406.